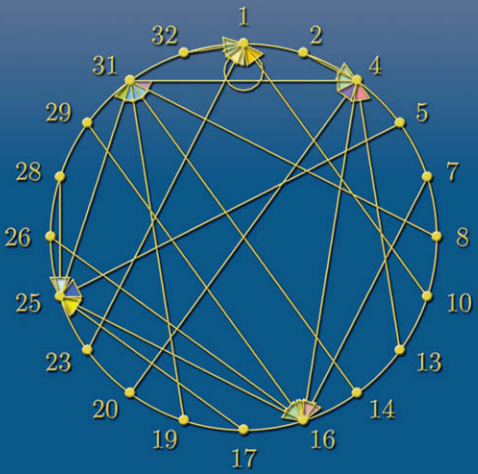
Wolfram Mathematica 7

GAP4

TEXTBOOKS in MATHEMATICS

Abstract Algebra

An Interactive Approach



William Paulsen

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

This page intentionally left blank

Abstract Algebra

An Interactive Approach

TEXTBOOKS in MATHEMATICS

Series Editor: Denny Gulick

PUBLISHED TITLES

ABSTRACT ALGEBRA: AN INTERACTIVE APPROACH
William Paulsen

COMPLEX VARIABLES: A PHYSICAL APPROACH WITH APPLICATIONS AND MATLAB®
Steven G. Krantz

ESSENTIALS OF TOPOLOGY WITH APPLICATIONS
Steven G. Krantz

INTRODUCTION TO ABSTRACT ALGEBRA
Jonathan D. H. Smith

INTRODUCTION TO MATHEMATICAL PROOFS: A TRANSITION
Charles E. Roberts, Jr.

LINEAR ALBEBRA: A FIRST COURSE WITH APPLICATIONS
Larry E. Knop

MATHEMATICAL AND EXPERIMENTAL MODELING OF PHYSICAL AND BIOLOGICAL PROCESSES
H. T. Banks and H. T. Tran

FORTHCOMING TITLES

ENCOUNTERS WITH CHAOS AND FRACTALS
Denny Gulick

TEXTBOOKS in MATHEMATICS

Abstract Algebra

An Interactive Approach

William Paulsen

Arkansas State University
Jonesboro, Arkansas, U.S.A.

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

Mathematica® and the *Mathematica* logo are registered trademarks of Wolfram Research, Inc. ("WRI" – www.wolfram.com and are used herein with WRI's permission. WRI did not participate in the creation of this work beyond the inclusion of the accompanying software, and offers it no endorsement beyond the inclusion of the accompanying software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2010 by Taylor and Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4200-9453-4 (Ebook)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

This page intentionally left blank

List of Figures

List of Tables

Preface

This textbook introduces a new approach to teaching an introductory course in abstract algebra. This text can be used for either an undergraduate level course, or a graduate level sequence. The undergraduate students would only cover the the basic material on groups and rings given in chapters 1–4 and 9–12. A graduate level sequence can be implemented by covering group theory in one semester (chapters 1–8), and covering rings and fields the second semester (chapters 9–15). Alternatively, one semester could cover part of the group theory chapters and part of ring theory, while the second semester covers the remainder of the book.

This text covers many graduate level topics that are not in most standard introductory abstract algebra courses. Some examples are semi-direct products (section 6.4), polycyclic groups (section 8.3), solving Rubik's Cube®-like puzzles (section 8.4), and Wedderburn's theorem (section 13.4). There are also some problem sequences that allow students to explore interesting topics in depth. For example, one sequence of problems outlines Fermat's two square theorem, while another finds a principal ideal domain that is not an Euclidean domain. Hopefully, these extra tidbits of information will satisfy the curiosity of the more advanced students.

What makes this book unique is the incorporation of technology into an abstract algebra course. Either *Mathematica*® or GAP (Groups, Algorithms, and Programming) can be used to give the students a hands-on experience to groups and rings. It is recommended to use at least one of these in the classroom. (GAP is totally free. See the section "*Mathematica* vs. GAP" for more information about both of these programs.) Every chapter includes several interactive problems that have the students use these programs to explore groups and rings. By doing these experiments, students can get a better grasp of the topic. However, there are plenty of non-interactive problems as well, so the instructor can choose not to force students into using these programs. The exception to this is in section 3.2, since the RSA encryption requires a computer program of some kind.

But in spite of the additional technology, this text is not short on rigor. There are still all of the classical proofs, although some of the harder proofs can be shortened with the added technology. For example, Abel's theorem is much easier to prove if we first assume that the 60-element group A_5 is simple, which *Mathematica* or GAP can verify in the classroom in less than a second. In fact, the added technology allows students to study larger groups, such as some of the Chevalley groups.

This text has many tools that will aid the students. There is a symbols table, so if a student sees an unfamiliar symbol, he can look up the description in this table, and see where this symbol is first defined. The answers to the odd-numbered problems are in the back, although the proofs are abbreviated. There is an extensive index that not only lists the relevant pages for a particular terminology, but also highlights the page where the term is first defined. A list of tables and figures allows students to find a multiplication table for a particular group or ring.

Acknowledgments

I am very grateful to Alexander Hulpke from Colorado State University for developing the GAP package "newrings.g" which allows GAP to work with finite rings. This package was specifically written for this textbook. Without this package, GAP would not be able to work with the examples that grace chapters 9–13. Other suggestions of his have proved to be invaluable.

I also must express my thanks to Shashi Kumar at the LaTeX help desk, who helped me with several different formatting issues throughout the text.

I also would like to express my appreciation to my wife Cynthia and my son Trevor for putting up with me during this past year, since this project ended up taking much more of my time than I first realized. They have been very patient with me and are looking forward to me finally being done.

This page intentionally left blank

About the Author

William Paulsen is a Professor of Mathematics at Arkansas State University. He has taught abstract algebra at both the undergraduate and graduate levels since 1997. He received his B.S. (summa cum laude), M.S., and Ph.D. degrees in mathematics at Washington University in St. Louis. He was on the winning team for the 45th William Lowell Putnam Mathematical Competition.

Dr. Paulsen has authored over 15 papers in abstract algebra and applied mathematics. Most of these papers make use of *Mathematica*®, including one which proves that Penrose tiles can be 3-colored, thus resolving a 30-year-old open problem posed by John H. Conway.

Dr. Paulsen has also programmed several new games and puzzles in Javascript and C++. One of these puzzles, Duelling Dimensions, is currently syndicated through Knight Features. Other puzzles and games are available on the Internet.

Dr. Paulsen lives in Harrisburg, Arkansas with his wife Cynthia, his son Trevor, and three pugs.

This page intentionally left blank

Symbol Description

x^{-1}	The inverse of the element x	4
(Mod n)	Modular arithmetic in base n	5
$x \equiv y$	x and y are in the same equivalence class	5, 72
$\mathbb{Z}$	The set of integers	10
GCD(m, n)	The greatest common divisor of m and n	12
$x \cdot y$	Group multiplication	15
e	Identity element of a group	16
$x \in G$	x is a member of the set or group G	16
$\|G\|$	Number of elements in a group or subgroup	17
Z_n	The group $\{0, 1, 2, \ldots, n-1\}$ using addition modulo n,	16
	or the ring of the same elements	264
Z_n^*	Numbers $< n$ coprime to n, with multiplication mod n	16
$\mathbb{Q}$	The group or field of rational numbers (fractions)	17
$\mathbb{Q}^*$	Non-zero rational numbers using multiplication	17
$\mathbb{R}$	The group or field of real numbers	17
$\mathbb{R}^*$	Non-zero real numbers using multiplication	86
x^n	x operated on itself n times	18
D_4	The group of symmetries of a square	22, 83
$\phi(n)$	Euler totient function	28
$\{\ldots \mid \ldots\}$	The set of elements ... such that ...	40
$H \cap K$	The intersection of H and K	40
$\bigcap_{H \in L} H$	The intersection of all sets in the collection L	41
$[S]$	Smallest subgroup containing the set S	41
$[x]$	Smallest subgroup containing the element x	43
xH	A left coset of the subgroup H	54
Hx	A right coset of the subgroup H	54
$H \backslash G$	The collection of right cosets of H in the group G	55
G/H	The collection of left cosets of H in the group G,	55
	or the quotient group of G with respect to H	71
$G \approx M$	The group G is isomorphic to M	80
Q	The quaternion group	84
$f : G \to M$	The function f maps elements of G to elements of M	86
Im(f)	The image (range) of the function f	90
$f^{-1}(x)$	The set of elements that map to x	90
$f^{-1}(H)$	The set of elements that map to an element of H	90
Ker(f)	The kernel of the homomorphism f, which is $f^{-1}(e)$	91

S_n	The symmetric group on n objects	109
$\left(\begin{smallmatrix}1&2&3&4\\2&3&4&1\end{smallmatrix}\right)$	Permutation notation	108
$n!$	n factorial $= 1 \cdot 2 \cdot 3 \cdots n$,	115
(12463)	Cycle notation	112
( )	The 0-cycle, the identity element of S_n	114
$\sigma(x)$	The signature function of the permutation x	118
A_n	The alternating group of permutations on n objects	119
$H \times K$	The direct product of the groups H and K	135
$\mathrm{Aut}(G)$	The group of automorphisms of the group G	152
$\mathrm{Inn}(G)$	The group of automorphisms of the group G	156
$\mathrm{Out}(G)$	The group of automorphisms of the group G	158
$H \ltimes_\phi N$	The semi-direct product of N with H through ϕ	161
D_n	The dihedral group with $2n$ elements	168
$Z(G)$	The center of the group G	176
$N_G(H)$	The normalizer of the subset H by the group G	179
$[H, K]$	The mutual commutator of the subgroups H and K	218
G'	The derived group of G, which is $[G, G]$	220
$\mathbb{H}$	The skew field of quaternions $a + bi + cj + dk$	251
$-x$	The additive inverse of x	252
nx	$x + x + x + \cdots + x$, n times	256
T_4	The smallest non-commutative ring	268
T_8	The smallest non-commutative ring with identity	268
$\overline{x}$	The conjugate of x	270, 330
R/I	The quotient ring of the ring R by the ideal I	281
$X * Y$	The product of two cosets in R/I	280
(S)	The smallest ideal containing the set S	283
(a)	The smallest ideal containing the element a	283
PID	Principal ideal domain	284, 373
$n\mathbb{Z}$	Multiples of n (also written as (n))	284
kZ_{kn}	Multiples of k in the ring Z_{kn}	287
$\mathbb{C}$	The field of complex numbers	306, 325
$\mathbb{Z}[x]$	The polynomials with integer coefficients	309
$K[x]$	The polynomials with coefficients in the ring K	309
$\left(\frac{x}{y}\right)$	The equivalence class of ordered pairs containing (x, y)	319
$R[a]$	Smallest ring containing the ring R and the element a	340
$\lvert z \rvert$	The absolute value of the complex number z	330
θ	Polar angle of a complex number	330
e^z	Complex exponential function	335
$\log(z)$	Complex logarithm function	336
x^z	Complex exponents	336
e_n	Principle n-th root of unity	333
$>_1$, $>_2$, $>_3$	Different ways of ordering the same ring,	345
ϕ_y	The homomorphism that evaluates a polynomial at y	354
UFD	Unique Factorization Domain	364

$R[x, y]$	The ring of polynomials of 2 variables over R	372
$\mu(x)$	The Euclidean valuation function	380
$N(a)$	The norm function on the ring $\mathbb{Z}(\sqrt{n})$	381
ϕ'	Restriction of a homomorphism to a smaller domain	392
$\Phi_n(x)$	The n-th cyclotomic polynomial	406
$GF(p^n)$	The field of order p^n	412
$K(S)$	The smallest field containing K and the set S	437
$K(a)$	The smallest field containing K and the element a	437
$\text{Irr}_F(a, x)$	The simplest polynomial in F with a as a root	440
$\mathbb{Q}(a, b)$	The smallest field containing $\mathbb{Q}$, a, and b	445
$\text{Gal}_F(K)$	The group of automorphisms of K which fix F	460
$\text{fix}(H)$	The field which is fixed by all automorphisms in H	467

This page intentionally left blank

Mathematica® *vs. GAP*

This textbook incorporates either GAP or *Mathematica* to help students visualize the important concepts of abstract algebra. It is recommended that one of the two programs be used with the book, but it is not necessary to have both. This section compares the two programs, and gives instructions for how to use these programs with the files on the included CD.

Mathematica is a symbolic manipulator package published by Wolfram Research, Inc. That is, it is a general purpose mathematical program used by scientists, engineers, and analysts. Its main feature that sets it apart from other symbolic manipulators is the graphics capabilities. In *Mathematica* 7.0, one can plot a 3-dimensional object, then use the mouse to rotate the object in three dimensions to see it from all possible angles.

GAP, on the other hand, has no graphics capabilities. It operates in a DOS window (or UNIX) and hence cannot display anything besides the letters that are typed on a keyboard. GAP stands for "Groups, Algorithms, and Programming," and is a system designed for computational work in abstract algebra. Hence, GAP is a natural program to use with an abstract algebra course.

Mathematica, however, was never designed to work problems involving abstract algebra. This is only possible via the two included packages "group.m" and "ring.m." The first of these is used for the first eight chapters of the text, while the other is used in the remaining chapters. Both files are in the "math" folder on the CD provided with this book. These are just two of the supporting files that are needed to use *Mathematica* or GAP with this textbook. These two files allow *Mathematica* to work with groups a fluently as GAP.

However, GAP has a big advantage over *Mathematica*—it is totally free. GAP is open source, which means that the source code is available to anyone who wishes to contribute to its vast library of abstract algebra operations. The current version of GAP can be downloaded from

`http://www.gap-system.org`

This textbook requires at least version 4.4.12 (December 2008). All of the outputs in this textbook use this version. Later versions may have slight differences, such as the order in which the terms appear, but this will not affect the functionality.

Mathematica is not free, but price information can be obtained from

`http://www.wolfram.com`

However, this book includes a 30-day *Mathematica* product trial. To download your trial, go to

`http://www.wolfram.com/books/resources`

and enter the license number below to be guided through the installation process.

`Trial license number: L3272-0591`

There is also a free *Mathematica Player* available from Wolfram, which will be able to open the notebooks provided with this textbook. However, one cannot execute any of the *Mathematica* commands with *Mathematica Player*. Those who are using GAP might consider downloading *Mathematica Player,* and directly opening the notebooks in the "math" folder on the CD to view some of the graphics that are unable to be displayed in GAP.

IMPORTANT: In order to use either GAP or *Mathematica* for this textbook, you will also need to install the supporting files into your computer. Simply put the CD provided into the computer, and the installation program should start running. If this program does not start automatically in any of the Windows versions, click on the "Start" icon, and select "Run." At this menu, select "Browse." and find the drive for the CD, and select the file "AbstractAlgebraSetup.exe." Hit "OK" to start the setup program running. Follow the instructions to install either the *Mathematica* or GAP supporting files, or both, onto the computer. Another option would be to copy the "math" and/or "gap" folders directly from the CD to the computer. This method will work in any operating system. Note that this only loads the supporting files, so you will also have to install *Mathematica* or GAP systems as well.

Once the supporting files have been installed, then one of the packages can be loaded into *Mathematica* with either of the two commands:

<< c:\math\group.m

<< c:\math\ring.m

This will only have to be done once in each *Mathematica* session.

Also in the supporting files are the 15 *Mathematica* notebooks "group01.nb" through "group08.nb," and "ring09.nb" through "ring15.nb" which correspond to the 15 chapters of the book. These notebooks allow a student to walk through the examples in the book, along with other similar examples. Included in these notebooks are all the theorems and proofs in the textbook.

The corresponding package for GAP, `textbook.g`, is in the "gap" folder on the included CD. Once the supporting software from the CD has been loaded to the computer, the GAP command

```
gap> Read("c:/gap/textbook.g");
```

will load the main package into the GAP session. As with *Mathematica*, this package *must* first be loaded into GAP before any other of the commands in this textbook will work.

Another of the supporting software files, "newrings.g" written by Alexander Hulpke, is used in chapters 9 through 15. This package is automatically loaded, if needed, when "textbook.g" is loaded. Future versions of GAP will have this library package incorporated into the program. Unlike *Mathematica*, GAP does not use notebooks, so all GAP commands shown in the textbook must be typed in manually. (Not even copying and pasting will work.)

Both of the programs are interactive systems. Every expression that one types into the computer is immediately evaluated, and the result is shown. This is known as a read-evaluate-print loop. For example, when GAP is first run, there will be a banner displayed, followed (eventually) by the GAP prompt

```
gap>
```

To enter an expression into GAP, simply end the expression with a semi-colon (;) and press the enter key.

```
gap> 3^90;
8727963568087712425891397479476727340041449
gap>
```

GAP echoes the answer on the next line, showing that GAP can handle numbers of enormous size. GAP then shows a new prompt to indicate that it is ready for the next problem. From now on, the textbook will not show this additional prompt.

Commands are entered into *Mathematica* a slightly different way. When the *Mathematica* program first opens, there are no prompts, but you can type anywhere into the "Untitled-1" window. Do not hit the enter key just yet.

3^90

Instead of ending with a semi-colon, hold down the Shift key while pressing the Enter key. Two things will happen: first a "In[1] :=" will appear in front of the expression you entered, and also the result will be displayed

In[1] := **3^90**
Out[1]:= 8 727 963 568 087 712 425 891 397 479 476 727 340 041 449

Mathematica will number all of the input and output statements, but the prompt does not appear until *after* some expression is entered. Because of this, the "In[n]:=" and "Out[n]:=" are not shown in the textbook.

Had we put a semi-colon in *Mathematica* before pressing the Shift-Enter, we would get a different effect. It computes the expression, but does not display the answer. For example, entering

a = 3^900;

in *Mathematica* will assign the variable a a 430 digit number, but will not display this number. To get this same effect in GAP, two semi-colons are needed.

```
gap> a := 3^900;;
```

Here is another difference between GAP and *Mathematica*. In *Mathematica*, the equal sign is used to assign an expression to a variable, whereas in GAP this is done with the `:=` combination, with no space between the colon and the equal sign.

In both programs, a variable is a sequence of letters and or digits, including at least one letter. *Mathematica* insists that the variable name start with a letter, which is a good practice to avoid confusion. Both programs are case sensitive, so a is a different variable than A. Keywords, such as `if` or `quit`, are not allowed as variables, but the list of keywords is too long to give here. None of the lower case letters are keywords, so we can safely use the 26 variables a through z.

Unlike GAP, *Mathematica* is able to have notebooks corresponding to each chapter. By clicking on "File" and then sliding down to "Open," one can locate one of the 15 notebooks with the .nb extension in the `c:\math` directory. When the notebooks are first opened, none of the "In[n] :=" or "Out[n] :=" will be present. This is because none of these commands has been executed in this particular session of *Mathematica*. The first command at the top will be the initialization, which will load either `group.m` or `ring.m`, which as we mentioned before must be done first. Click on the bold-face command to have the cursor on this command (it doesn't have to be at the end) and press Shift-Enter. Now the "In[1] := " will appear, showing that this command has been executed. All other bold-face commands can be executed the same way. It is suggested that this be done in the order that they appear, but there is nothing to prevent executing the statements in any order, or executing a statement more than once. The "In[n] :=" and "Out[n] :=" will show which commands have been run and in what order. Just because there is an output displayed for some input does not mean that this input has been executed. For example, if the notebook displays

a = 3^90
8 727 963 568 087 712 425 891 397 479 476 727 340 041 449

and there is no "In[n] :=" in front of the line, then the value of a will still be undefined *even though the output is already displayed.* This output is from a previous session of *Mathematica*, and all variables are reset at the beginning of each session. So for a to be given the value of 3^{90}, this must be re-evaluated using the Shift-Enter. *Mathematica* will then evaluate 3^{90} and of course come up with the same answer, but this time a "Out[n] :=" will appear in front of the answer to show that it has been executed.

Mathematica does not automatically expand an expression, although it might rearrange the factors and terms.

(x^2+3x–1)(x^2–2x+4)
$(4 - 2x + x^2)(-1 + 3x + x^2)$

Because we have not yet assigned a value to x, *Mathematica* assumes that it is an indeterminate, so that it expresses the answer in terms of x. Also note that *Mathematica* assumes that a number and letter next to each other are to be multiplied together. In GAP, we must explicitly use the `*` for every multiplication.

```
gap> (x^2+3*x-1)*(x^2-2*x+4);
Variable: 'x' must have a value
```

This time, get an error message, since GAP has not been told what x is yet. Unlike *Mathematica*, GAP must have something assigned to a variable in order to use it. If we want x to be an unknown quantity, or indeterminate, we must assign to the variable x an indeterminate form. Basically, this tells GAP that x is to be treated as an unknown quantity, but of a certain type. In this case, we will suppose that x is an unknown rational number. (GAP is not able to work with general real numbers—more about this later.) While we are at it, we can tell GAP how this variable is to be displayed.

```
gap> x := Indeterminate(Rationals,"x");
x
gap> (x^2+3*x-1)*(x^2-2*x+4);
x^4+x^3-3*x^2+14*x-4
```

GAP will automatically expand the expression. In order to do this in *Mathematica*, the **Expand** function is necessary.

Expand[%]
$-4 + 14x - 3x^2 + x^3 + x^4$
Factor[%]
$(4 - 2x + x^2)(-1 + 3x + x^2)$

Note that *Mathematica* uses the percent sign (%) as an abbreviation for the last output. The corresponding GAP abbreviation is `last`.

```
gap> Factors(last);
[ x^2-2*x+4, x^2+3*x-1 ]
```

Here is another syntax difference between GAP and *Mathematica*: GAP uses parentheses for functions, as the standard notation, but *Mathematica* uses square brackets for functions. GAP mainly uses the square brackets for lists, so the output shows a list of the factors.

Note that we defined x to be a rational variable, not a real variable. The truth is, GAP never works with real numbers or decimals. Since GAP is only designed for working with groups, rings, and other similar objects, there is no need for decimals. This means that all calculations done in GAP are *exact*. Most calculations in *Mathematica* are also exact, but you do have the option of finding a decimal approximation. For example, the first 50 digits of $\sqrt{2}$ are

N[Sqrt[2],50]
1.4142135623730950488016887242096980785696718753769

We get a surprise when we try to find $\sqrt{2}$ in GAP:

```
gap> Sqrt(2);
E(8)-E(8)^3;
```

GAP puts the answer in terms of a number e_8, which we will cover in section 11.3. Other common irrational numbers, such as π, cannot be entered into GAP at all! This is only because of the specialized nature of the GAP program.

Both GAP and *Mathematica* will point out any mistakes in the input line. For example, if one types

```
gap> y := Indeterminate(Integers, "y");
y
gap> (y+2)(y+4);
Syntax error: ; expected
(y+2)(y+4);
     ^
```

GAP will realize a mistake, and point to the error with an arrow (^). GAP will try to read your mind as to your intentions, and apparently GAP thought that we were trying to input two expressions on the same line, separated by a semi-colon. But in fact, we forgot the multiplication symbol. Rather than retyping the line, we can press the up arrow key (or Ctrl-P) and the last line will be redisplayed with the error. We then can use the arrow keys (or Ctrl-B and Ctrl-F) to get to the erroneous location and fix the problem.

```
gap> (y+2)*(y+4);
y^2+6*y+8
```

Occasionally, GAP will encounter an error that it cannot handle, and enter into a *break loop*. After a fairly long error message, a special prompt `brk>` appears. This prompt is very useful for debugging the program to find just where the error occurred and why, but for our purposes the best thing to do is to enter `quit;` at the break prompt, and we will return to the place before the error. Entering `quit;` at the `gap>` prompt will exit the program.

The most common error message of this type is the "no method found!" error, which is at first rather cryptic.

```
gap> (4 = 3) * 2;
Error, no method found!
For debugging hints type ?Recovery from NoMethodFound
Error, no 1st choice method found for 'PROD' on 2 arguments
called from <function>( <arguments> ) called from read-eval-loop
Entering break read-eval-print loop ...
you can 'quit;' to quit to outer loop, or
you can 'return;' to continue
brk>
```

This error means that GAP tried to perform some operation on an object whose type did not match the operation. In this case, GAP simplified the expression in the parentheses first. A standard equal sign is used to test whether two objects are equal, and since 4 and 3 are obviously not equal, this simplified to `false`. But then GAP tried to calculate `false * 2`, and looked through the libraries to see if there is any method to multiply an integer with `false`. Obviously, there is no such method, hence the error message.

The way to recover from this is to first type in `quit;` to get out of the break loop, then fix the mistake.

```
brk> quit;
gap> (4 + 3) * 2;
14
```

The same typo also produces an error in *Mathematica*, but for a different reason.

(4 = 3) * 2
6

Mathematica returns an answer, but also displays a strange message,

"Set::setraw : Cannot assign to raw object 4. ≫"

in a separate Messages window. Because the equal sign in *Mathematica* is used to assign a value to a variable, *Mathematica* thinks we are trying to assign the value 3 to the number 4, which of course cannot be done. But besides this, this value of 3 is multiplied by 2 to get the answer displayed.

Ironically, had we used a double equal sign, the *Mathematica* command would not have produced an error.

(4 == 3) * 2
2 False

The double equal sign is used in *Mathematica* to test if two expressions are equal. Unlike GAP, *Mathematica* sees no problem in symbolically multiplying **False** with an integer.

Other features of the two programs will be introduced in the textbook as the need arises. With a little practice, you will find both programs are relatively easy to use.

This page intentionally left blank

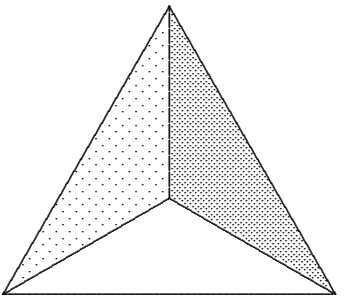
Chapter 1

Understanding the Group Concept

1.1 Introduction to Groups

To help introduce us to the concept of groups, let us meet a triangle whose dance steps give us the first example of a group. Terry the triangle is a simple looking three-colored triangle that appears by the *Mathematica*® command

ShowTerry



Terry can perform the dance steps listed in table 1.1. Although *Mathematica* animates these dance steps, one can understand the six steps without *Mathematica* by observing scenes in figure 1.1, taken from the animation close to the completion of each step.

Terry can combine these dance steps to form a dance routine. But in any routine, the ending position of the triangle is the same as that of performing just one dance step. Thus, when the triangle gets "lazy," it can perform just one dance step instead of several. For example, a **FlipRt** followed by a **Spin**

TABLE 1.1: Terry's dance steps

RotRt	rotate clockwise 120 degrees.
RotLft	rotate counterclockwise 120 degrees.
Spin	spins in three dimensions, keeping the top fixed.
FlipRt	flips over the right shoulder.
FlipLft	flips over the left shoulder.
Stay	does nothing.

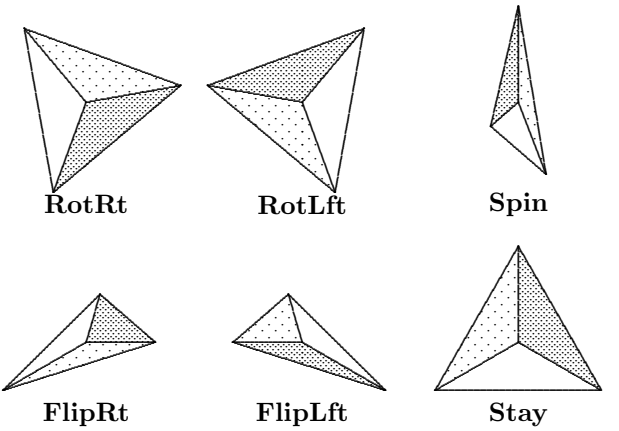


FIGURE 1.1: Scenes from Terry's animated dance steps

TABLE 1.2: Multiplication table for Terry's dance steps

	Stay	FlipRt	RotRt	FlipLft	RotLft	Spin
Stay	Stay	FlipRt	RotRt	FlipLft	RotLft	Spin
FlipRt	FlipRt	Stay	FlipLft	RotRt	Spin	RotLft
RotRt	RotRt	Spin	RotLft	FlipRt	Stay	FlipLft
FlipLft	FlipLft	RotLft	Spin	Stay	FlipRt	RotRt
RotLft	RotLft	FlipLft	Stay	Spin	RotRt	FlipRt
Spin	Spin	RotRt	FlipRt	RotLft	FlipLft	Stay

puts Terry in the same position as a **RotLft**. These dance steps are combined using the "multiplication table" in table 1.2.

The *Mathematica* commands that generated this table are

InitTerry;
MultTable[{**Stay, FlipRt, RotRt, FlipLft, RotLft, Spin**}]

whereas the corresponding GAP commands are

```
gap> Read("c:/gap/textbook.g");
gap> InitTerry();
[ Stay, FlipRt, RotRt, FlipLft, RotLft, Spin ]
gap> MultTable(Terry);
*       |Stay     FlipRt   RotRt    FlipLft  RotLft   Spin
--------+-----------------------------------------------------
Stay    |Stay     FlipRt   RotRt    FlipLft  RotLft   Spin
FlipRt  |FlipRt   Stay     FlipLft  RotRt    Spin     RotLft
RotRt   |RotRt    Spin     RotLft   FlipRt   Stay     FlipLft
FlipLft |FlipLft  RotLft   Spin     Stay     FlipRt   RotRt
RotLft  |RotLft   FlipLft  Stay     Spin     RotRt    FlipRt
Spin    |Spin     RotRt    FlipRt   RotLft   FlipLft  Stay
```

which produce an ASCII facsimile of the table. In both these tables, the first dance steps are on the left, and the second dance steps are on the top, so one can use the table to see that **FlipRt** · **Spin** = **RotLft**.

We can notice several things from these dance steps:

1. The *order* in which the dance steps are performed are important. For example, **Spin** · **FlipRt** ≠ **FlipRt** · **Spin**.

2. The combination of any two dance steps is equivalent to one of the six dance steps. In other words, there are no "holes" in table 1.2.

3. The order in which a dance routine is simplified does not matter. That is,
$$x \cdot (y \cdot z) = (x \cdot y) \cdot z$$
where x, y, and z represent three dance steps.

4. Any dance step combined with **Stay** yields the same dance step. This is apparent by looking at the row and column corresponding to **Stay** in table 1.2.

5. Every dance step has another dance step that "undoes" it. That is, for every x there is a y such that $x \cdot y =$ **Stay**. For example, the step that undoes **RotRt** is **RotLft**.

We will use the following mathematical terminology to express each of these properties:

1. The dance steps are not *commutative.*

2. The dance steps are *closed* under multiplication.

3. The dance steps are *associative.*

4. There is an *identity* dance step.

5. Every dance step has an *inverse.*

With just these properties, we are able to prove the following.

PROPOSITION 1.1
If y is an inverse of x, then x is the only inverse of y.

PROOF Let z be any inverse of y. Consider the product $x \cdot y \cdot z$. According to the associative property,

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z.$$

On the left side, we see that $y \cdot z$ is an identity element, so $x \cdot (y \cdot z) = x$. But on the right side, we find that $x \cdot y$ is an identity element, so $(x \cdot y) \cdot z = z$. Therefore, $x = z$, and so x is the only inverse of y. □

Notice that we did not yet assume that the inverses are unique, or even that there is only one identity element. However, these facts immediately follow from proposition 1.1. (See problems 1.8 and 1.9.)

DEFINITION 1.1 We use the notation x^{-1} for the unique inverse of the element x.

Proposition 1.1 can now be expressed simply as $(x^{-1})^{-1} = x$. This raises the question as to whether other familiar exponential properties hold. For example, does $(x \cdot y)^{-1}$ always equal $x^{-1} \cdot y^{-1}$?

```
gap> (FlipRt*Spin)^-1 = (FlipRt^-1)*(Spin^-1);
false
```

Apparently $(x \cdot y)^{-1}$ is not always equal to $x^{-1} \cdot y^{-1}$. Yet it is not hard to determine the correct way to simplify $(x \cdot y)^{-1}$.

PROPOSITION 1.2

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

PROOF Since the inverse $(x \cdot y)^{-1}$ is the unique dance step z such that

$$(x \cdot y) \cdot z = \mathbf{Stay},$$

it suffices to show that $y^{-1} \cdot x^{-1}$ has this property. We see that

$$(x \cdot y) \cdot (y^{-1} \cdot x^{-1}) = x \cdot (y \cdot y^{-1}) \cdot x^{-1} = x \cdot \mathbf{Stay} \cdot x^{-1} = x \cdot x^{-1} = \mathbf{Stay}.$$

So $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$. □

Another pattern of the multiplication table of the dance steps is that each row and each column in the interior part of the table contain all six dance steps. For example, **RotRt** appears only once in the row beginning with **Spin**. That is, there is only one solution to $\mathbf{Spin} \cdot x = \mathbf{RotRt}$. We can show why this pattern holds in general using inverses.

PROPOSITION 1.3
If a and b are given, then there exists a unique x such that

$$a \cdot x = b.$$

PROOF Suppose that there is an x such that $a \cdot x = b$. We can multiply both sides of the equation on the *left* by a^{-1} to give us

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

Then

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

$$\mathbf{Stay} \cdot x = a^{-1} \cdot b.$$

So

$$x = a^{-1} \cdot b.$$

Thus, if there is a solution, this must be the unique solution $x = a^{-1} \cdot b$. Let us check that this is indeed a solution.

$$a \cdot (a^{-1} \cdot b) = (a \cdot a^{-1}) \cdot b = \mathbf{Stay} \cdot b = b.$$

Thus, there is only one solution to the equation, namely $a^{-1} \cdot b$. □

This last proposition, when combined with problem 1.10, shows that the interior of the multiplication table forms a "Latin square." A Latin square is a formation in which every row and every column contain each item once and only once. The Latin square property is easy to check visually.

Even though there are very few of Terry's dance steps, we already can see some of the patterns that can appear when we consider the multiplication of these dance steps. In the next section, we will consider another operation that has many of the same patterns.

1.2 Modular Arithmetic

We have already seen that one operation, namely the combination of Terry's dance steps, turns out to have some interesting properties such as the Latin square property. In this section we will find some other operations that have this same property. These will involve the *modulus* of a number, and so we must study the arithmetic on numbers modulo n.

The simple definition of the modulus of a number is the last digit of the number when written in base n. We can also consider the modulus of a number to be the *remainder* when that number is divided by n. Two numbers are considered *equivalent modulo* n if the modulus of the numbers are the same. The official definition is as follows.

DEFINITION 1.2

$$x \equiv y \ (\text{Mod } n)$$

if, and only if, there is an integer k such that

$$(x - y) = k\,n.$$

We first consider adding numbers together modulo 10. That is, after each addition, we only consider the last digit of the result. The command

DefSumMod[10]

loads this new type of arithmetic into *Mathematica*. The period is then used to add together to numbers from the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ modulo 10. Although it seems strange to use the period instead of the plus sign, for consistency *Mathematica* always uses the period for the operator, whatever operation that operator performs. GAP can also be used to explore addition modulo 10.

```
gap> (6 + 7) mod 10;
3
gap> (9 + 8) mod 10 = (8 + 9) mod 10;
true
```

The table for this operation on the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is given by

```
gap> MultTable([0..9]);

+|0  1  2  3  4  5  6  7  8  9
-+--------------------------
0|0  1  2  3  4  5  6  7  8  9
1|1  2  3  4  5  6  7  8  9  0
2|2  3  4  5  6  7  8  9  0  1
3|3  4  5  6  7  8  9  0  1  2
4|4  5  6  7  8  9  0  1  2  3
5|5  6  7  8  9  0  1  2  3  4
6|6  7  8  9  0  1  2  3  4  5
7|7  8  9  0  1  2  3  4  5  6
8|8  9  0  1  2  3  4  5  6  7
9|9  0  1  2  3  4  5  6  7  8
```

The corresponding *Mathematica* command would be

DefSumMod[10]
MultTable[{0,1,2,3,4,5,6,7,8,9}]

Notice that we still call this a "multiplication table" even though the operation is closer to addition. Only in GAP can we use the abbreviation `[0..9]` for the list `[0,1,2,3,4,5,6,7,8,9]`. Also, GAP uses context to determine that we are to add modulo 10.

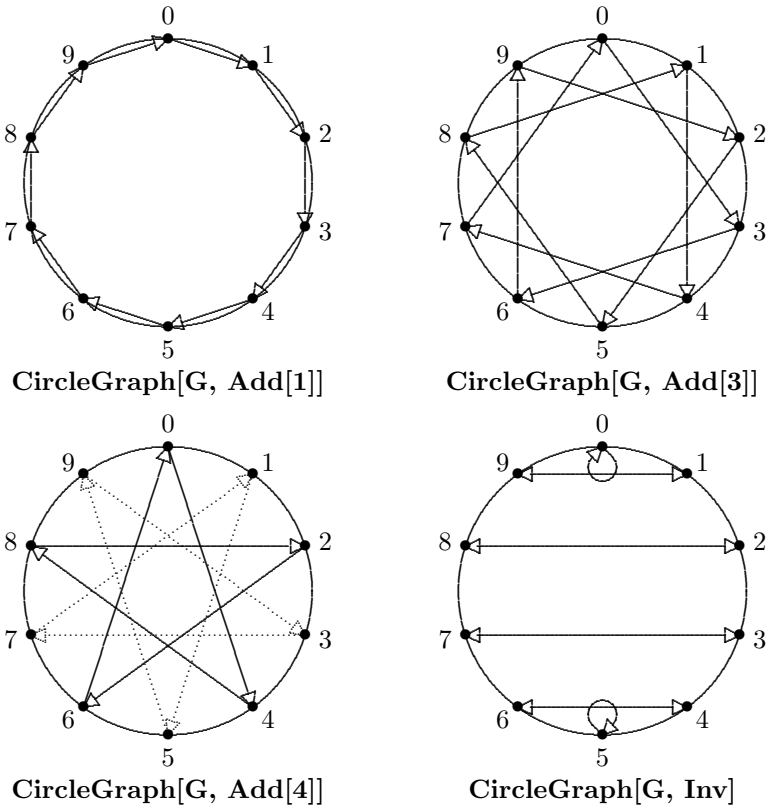
Having the table for addition modulo 10, we are able to establish the following properties:

1. For any two numbers x and y in $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, $x \cdot y$ is in the set. (Recall that we are using the dot to indicate the operation, regardless of what that operation is. In this example, the operation is addition modulo 10.)

2. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for any x, y, and z.
3. $x \cdot 0 = x$ and $0 \cdot x = x$ for all x.
4. For any x, there is a y such that $x \cdot y = 0$.
5. For any x and y, $x \cdot y = y \cdot x$.

This operation can also be pictured by means of circular graphs. The *Mathematica* command

G = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}
CircleGraph[G, Add[1]]



CircleGraph[G, Add[1]] **CircleGraph[G, Add[3]]**

CircleGraph[G, Add[4]] **CircleGraph[G, Inv]**

FIGURE 1.2: Circle graphs for (Mod 10) arithmetic

gives us the first picture in figure 1.2, which draws an arrow from each point to the point given by "adding 1 modulo 10." Figure 1.2 also shows what

happens if we replace the 1 with 3 or 4. We get different looking graphs, but all with the same amount of symmetry. The *Mathematica* command

CircleGraph[G, Add[1],Add[2],Add[3],Add[4],Add[5]]

combines several of these circular graphs together, each drawn in a different color. The last picture in figure 1.2 shows the additive inverse of each digit. This was created with the command

CircleGraph[G, Inv]

Of course, we could do these same experiments by considering addition modulo n with any other base as well as $n = 10$. The patterns formed by the circular graphs are very similar. But we can also consider the operation of *multiplying* modulo n. The *Mathematica* command

DefMultMod[7]

defines the period to be multiplication modulo 7. The multiplication table of this new operation has similar properties as the table of dance steps for the triangle, especially if we removed the 0 and only considered the digits $\{1, 2, 3, 4, 5, 6\}$. The identity element is 1, and each of the numbers has an inverse. The GAP command

```
gap> MultTable([1..6]);

*|1  2  3  4  5  6
-+---------------
1|1  2  3  4  5  6
2|2  4  6  1  3  5
3|3  6  2  5  1  4
4|4  1  5  2  6  3
5|5  3  1  6  4  2
6|6  5  4  3  2  1
```

displays the multiplication table.

But when we try using a different base, we get a surprise. To display the multiplication table for (Mod 10) arithmetic, we can either use the GAP command `MultTable([1..9])`, or the *Mathematica* commands

DefMultMod[10]
MultTable[{0,1,2,3,4,5,6,7,8,9}]

to produce a table similar to table 1.3. We find several rows that do not contain any 1's. These rows indicate the numbers without inverses modulo 10. Only $1, 3, 7$, and 9 have inverses. If we try this using 15 instead of 10, we find only $1, 2, 4, 7, 8, 11, 13$, and 14 have inverses.

But what if we consider the multiplication table of just those numbers that have inverses modulo 15? We can use either the *Mathematica* commands

DefMultMod[15]
MultTable[{1, 2, 4, 7, 8, 11, 13, 14}]

TABLE 1.3: Multiplication (Mod 10)

	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9
2	0	2	4	6	8	0	2	4	6	8
3	0	3	6	9	2	5	8	1	4	7
4	0	4	8	2	6	0	4	8	2	6
5	0	5	0	5	0	5	0	5	0	5
6	0	6	2	8	4	0	6	2	8	4
7	0	7	4	1	8	5	2	9	6	3
8	0	8	6	4	2	0	8	6	4	2
9	0	9	8	7	6	5	4	3	2	1

TABLE 1.4: Invertible elements (Mod 15)

	1	2	4	7	8	11	13	14
1	1	2	4	7	8	11	13	14
2	2	4	8	14	1	7	11	13
4	4	8	1	13	2	14	7	11
7	7	14	13	4	11	2	1	8
8	8	1	2	11	4	13	14	7
11	11	7	14	2	13	1	8	4
13	13	11	7	1	14	8	4	2
14	14	13	11	8	7	4	2	1

or the GAP command `MultTable([1,2,4,7,8,11,13,14]);` to produce table 1.4. Once again, many of the same patterns are found that were in for Terry's multiplication, namely:

1. For any two numbers x and y in $\{1, 2, 4, 7, 8, 11, 13, 14\}$, $x \cdot y$ is in that set.

2. $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for any x, y, and z.

3. $x \cdot 1 = x$ and $1 \cdot x = x$ for all x.

4. For any x, there is a y such that $x \cdot y = 1$.

5. For any x and y, $x \cdot y = y \cdot x$.

We can generalize these patterns to multiplication modulo n for any n.

PROPOSITION 1.4

For n a positive integer greater than 1, let the dot $(\cdot)$ denote multiplication modulo n. Let G be the set of all non-negative numbers less than n that have inverses modulo n. Then the set G has the following properties:

1. *For any two numbers x and y in G, $x \cdot y$ is in G.*
2. *$(x \cdot y) \cdot z = x \cdot (y \cdot z)$ for any x, y, and z.*
3. *$x \cdot 1 = 1 \cdot x = x$ for all x.*
4. *For any x that is in G, there is a y in G such that $x \cdot y = 1$.*
5. *For any x and y, $x \cdot y = y \cdot x$.*

PROOF Properties 2, 3, and 5 come from the properties of standard multiplication.

Property 1 comes from proposition 1.2. If x and y are both invertible, then $y^{-1} \cdot x^{-1}$ is an inverse of $x \cdot y$, and so $x \cdot y$ is invertible modulo n.

Property 4 seems obvious, since if x is invertible modulo n, we let $y = x^{-1}$ making $x \cdot y = 1$. But we must check that y is also invertible, which it is since $y^{-1} = x$. ∎

Of course, this does not tell us *which* of the numbers less then n have inverses modulo n. To answer this question, we must first explore the prime factorizations of numbers, and properties that this imposes onto the integers.

1.3 Prime Factorizations

In this section we will explore the basic properties of integers stemming from the prime factorizations. We will denote the set of all integers,

$$\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$$

by the stylized letter $\mathbb{Z}$. This notation comes from the German word for number, *Zahl.* Many of the properties of factorizations refer only to positive integers, which are denoted $\mathbb{Z}^+$.

We define a *prime* as an integer that has only two positive factors: 1 and itself. This definition actually allows negative numbers, such as -5, to be prime. Although this may seem to be a nonstandard definition, it agrees with the generalized definition of primes defined in chapters 10 and 12. The numbers 1 and -1 are not considered to be prime. The familiar property of primes is that any integer greater than 1 can be uniquely factored into a

product of positive primes. The uniqueness aspect of this statement will be proven in chapter 12, in a much more generalized context. We will begin by proving that every large number has at least one prime factor.

LEMMA 1.1
Every number greater than 1 has a prime factor.

PROOF Suppose that some number greater than 1 does not have a prime factor. Then there is a smallest such number, called n. Then n is not prime, otherwise n would have a prime factor. Then by definition, n must have a positive divisor besides 1 and n, say m. Since $1 < m < n$, and n was the smallest number greater than 1 without a prime factor, m must have a prime factor, say p. Then p is also a prime factor of n, so we have a contradiction. Therefore, every number greater than 1 has a prime factor. □

The proof of lemma 1.1 introduces an important strategy in proofs. Notice that to prove that every number greater than 1 had a prime factor, we assumed just the opposite. It was as if we admitted defeat from the very beginning! Yet from this we were able to reach a conclusion that was absurd—a number without a prime factor that did have a prime factor. This strategy is known as *reductio ad absurdum*, which is Latin for "reduce to the absurd." We assume what we are trying to prove is actually false, and proceed logically until we reach a contradiction. The only explanation would be that the assumption was wrong, which proves the original statement.

In problem 1.33, you will be asked to use lemma 1.1 to prove that every positive integer can be written as a product of primes. The *Mathematica* command for finding the prime factorization of an integer is

FactorInteger[420]
$\{\{2,2\},\{3,1\},\{5,1\},\{7,1\}\}$

whereas the gap command is

```
gap> FactorsInt(420);
[ 2, 2, 3, 5, 7 ]
```

Mathematica lists the primes, along with how many times that prime divides the number. GAP, on the other hand, can list the same prime several times. As long as the integers are less than about 40 digits long, neither program should have any trouble factoring them. However, integer factorization is a difficult problem even with modern technology. For both programs, the amount of time required is proportional to the square root of the second largest prime in the factorization. [14, p. 133]

The prime factorizations lead to an important question. Is there a largest prime number? The Greek mathematician Euclid answered this question using *reductio ad absurdum* in the third century B.C. [11, p. 183]

THEOREM 1.1: Euclid's Prime Number Theorem
There are an infinite number of primes.

PROOF Suppose that there are only a finite number of prime numbers. Label these prime numbers

$$p_1 = 2, \quad p_2 = 3, \quad p_3 = 5, \quad \ldots, \quad p_n.$$

Now consider the number

$$m = (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdots p_n) + 1$$

This number is odd, so it cannot be divisible by 2. Likewise, m is one more than a multiple of 3, so it is not divisible by 3. In this way we see that m is not divisible by any of the prime numbers. But this is ridiculous, since m must have a prime factor by lemma 1.1. Thus, the original assumption that there is a largest prime number is false, so there are an infinite number of prime numbers. □

We define the *greatest common divisor* (GCD) of two numbers to be the largest integer that divides both of the numbers. If the greatest common divisor is 1, this means that there are no prime factors in common. We say the numbers are *coprime* in this case. We can use *Mathematica*'s **GCD** function or GAP's `GcdInt` function to quickly test whether two numbers are coprime without having to factor them.

GCD[138153809229555633320990299469,
145730407810127891189961221324529]

```
gap> GcdInt(138153809229555633320990299469,
> 145730407810127891189961221324529);
1
```

There is an important property of the greatest common divisor, given in the following theorem.

THEOREM 1.2: The Greatest Common Divisor Theorem
Given two positive integers x and y, the greatest common divisor of x and y is the smallest positive integer which can be expressed in the form

$$ux + vy$$

with u and v being integers.

PROOF Let A denote the set of all positive numbers that can be expressed in the form $u \cdot x + v \cdot y$. Note that both x and y can be written in the form

$u \cdot x + v \cdot y$, so we can consider the smallest positive number n that can be written in the form $u \cdot x + v \cdot y$. Note that $\mathrm{GCD}(x, y)$ is a factor of both x and y, so $\mathrm{GCD}(x, y)$ must be a factor of n.

Next, consider the number

$$k \equiv x \ (\mathrm{Mod}\ n) \quad \text{with} \quad 0 \leq k < n.$$

Then $k = x + nr$ for some number r. But $n = ux + vy$ for some numbers u and v. Thus,

$$k = x + (ux + vy)r = (1 + ru)x + (rv)y,$$

so k is in A. But since n is the smallest positive integer in A, k cannot be equivalent (Mod n) to any number less than n, other than 0. Thus,

$$x \equiv 0 \ (\mathrm{Mod}\ n).$$

Therefore, n is a divisor of x. By a similar reasoning, n is also a divisor of y. Thus, n is a common divisor of x and y, and since the $\mathrm{GCD}(x, y)$ is in turn a divisor of n, n must be equal to $\mathrm{GCD}(x, y)$. □

We can find the numbers u and v from the greatest common divisor theorem (1.2) using either *Mathematica* or GAP. The *Mathematica* command

ExtendedGCD[105, 196]
{7, {-13, 7}}

gives the GCD to be 7, and also says that $u = -13$ and $v = 7$ will satisfy theorem 1.2, so $(-13) \cdot 105 + 7 \cdot 196 = 7$. The corresponding GAP commands

```
gap> GcdInt(105,196);
7
gap> Gcdex(105,196);
rec( gcd := 7, coeff1 := -13, coeff2 := 7, coeff3 := 28,
  coeff4 := -15 )
```

give even more information. The u and v are given by `coeff1` and `coeff2`, giving the same result as *Mathematica*, but GAP gives two more numbers that can be added (or subtracted) to u and v to produce different answers. Thus, $u = -13 + 28 = 15$ and $v = 7 + (-15) = -8$ is another solution.

In the last section we found that the invertible elements modulo n had many of the properties of Terry's dance steps. We now can determine which numbers less that n have a multiplicative inverse modulo n.

PROPOSITION 1.5

Let n be in $\mathbb{Z}^+$. Then for x between 0 and $n-1$, x has a multiplicative inverse modulo n if, and only if, x is coprime to n.

PROOF If x and n are not coprime, then there is a common prime factor p. In order for x to have a multiplicative inverse, there must be a y such that

$$x \cdot y \equiv 1 \pmod{n}$$

But this means that $xy = 1 + wn$ for some w. This is impossible, since xy is a multiple of p, but $1 + wn$ is one more than a multiple of p.

Now suppose that x and n are coprime. By the greatest common divisor theorem (1.2), there are u and v in $\mathbb{Z}$ such that

$$ux + vn = \text{GCD}(x, n) = 1.$$

But then

$$ux = 1 + (-v)n,$$

and so $u \cdot x \equiv 1 \pmod{n}$. Hence, u is a multiplicative inverse of x. □

There is another property of modular arithmetic involving coprime numbers that will be used often throughout the book, known to the ancient Chinese.

THEOREM 1.3: The Chinese Remainder Theorem
If u and v in $\mathbb{Z}^+$ are coprime, then given any x and y in $\mathbb{Z}$, there is a unique k in $\mathbb{Z}$ such that

$$0 \leq k < u \cdot v,$$

$$k \equiv x \pmod{u},$$

and

$$k \equiv y \pmod{v}.$$

PROOF Ironically, the way that we will show that there is such a number is to show that there cannot be more than one such number!

Suppose we have two different numbers, k and q, which satisfy the above conditions. Then

$$k - q \equiv 0 \pmod{u} \qquad \text{and} \qquad k - q \equiv 0 \pmod{v}.$$

Thus, $k - q$ must be a multiple of both u and v. But since u and v are coprime, the least common multiple of u and v is $u \cdot v$. Thus, $k - q$ is a multiple of $u \cdot v$.

However, both k and q are less then $u \cdot v$. So the only way this is possible is for $k - q = 0$, which contradicts our assumption that k and q were distinct solutions.

Thus, we have shown that there cannot be more than one value for k. But how does that help us prove that there must be such a k? Notice that for any number k,

$$k \pmod{u}$$

can have u possible values, from 0 to $u-1$. Also,

$$k \text{ (Mod } v)$$

can have v possible values, from 0 to $v-1$. Thus, for any k, there are only $u \cdot v$ possible values for the ordered pair

$$(k \text{ (Mod } u), k \text{ (Mod } v)).$$

What is shown above is that no two values of k between 0 and $u \cdot v - 1$ can give the same ordered pair. But there are exactly $u \cdot v$ such values of k.

Imagine having $u \cdot v$ "pigeonholes" labeled by these ordered pairs. If one has $u \cdot v$ pigeons and $u \cdot v$ pigeonholes, and each pigeon goes into a pigeonhole with no two pigeons going into the same hole, then every hole must be filled!

In the same way, since each of the $u \cdot v$ possible values of k produces one of the $u \cdot v$ possible ordered pairs, and no two k's can produce the same ordered pair, each ordered pair must be produced by some (unique) value of k. And this is what we wanted to prove. □

This proof introduced a second technique to prove theorems, called the *pigeonhole principle.* Whenever we have a mapping from n objects into n other objects, and there are no duplications, then there must be a one-to-one correspondence between the two sets of objects. This is an important principle that we will use several times throughout this book.

Ironically, using the pigeonhole principle does not give us a way (short of trial and error) of finding the value of k. However, there is a GAP command that finds k given the 2 sets $\{u, v\}$ and $\{x, y\}$:

```
gap> ChineseRem([125,81],[23,17]);
4148
gap> 4148 mod 125;
23
gap> 4148 mod 81;
17
```

1.4 The Definition of a Group

We are now ready to try to generalize the examples we have studied. We will define a *group* abstractly using only the properties that all of our examples had in common.

DEFINITION 1.3 A *group* is a set G together with an operation $(\cdot)$ such that the following four properties hold:

1. (closure) For any x and y in G, $x \cdot y$ is in G.

2. (identity) There exists a member e in G which has the property that $e \cdot x = x \cdot e = x$ for all x in G.

3. (inverse) For every x in G, there exists a y in G, called the *inverse* of x, such that $x \cdot y = e$.

4. (associative law) For any a, b, and c in G, then $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

Terry's dance steps give us the first example of a group, more commonly referred to as the group of symmetries of a triangle, D_3.

The members of the group, whether they are numbers, dance steps, or even ordered pairs, are called the *elements* of the group. The element e that satisfies property 2 is called the *identity element* of the group.

The mathematical notation for an element x to be in a group G is

$$x \in G.$$

Since propositions 1.1, 1.2, and 1.3 used only these four properties, the proofs are valid for all groups, using the identity element e in place of the dance step `Stay`.

Other examples of groups come from modular arithmetic. For n in $\mathbb{Z}^+$, we considered the elements

$$\{0, 1, 2, ..., n-1\},$$

with the operator $(\cdot)$ being the sum modulo n. This group will be denoted Z_n.

We also considered having the operator $(\cdot)$ denote the product modulo n, and considered only the set of numbers less than n that are coprime to n. Proposition 1.4 shows that this set also has the four properties of groups. We will refer to this group by Z_n^*.

The groups Z_n and Z_n^* had a fifth property—the multiplication tables were symmetric about the northwest to southeast diagonal. Not all groups have this property, but those that do are important enough to give such groups a special name.

DEFINITION 1.4 A group G is *abelian* (or *commutative*) if $x \cdot y = y \cdot x$ for all $x, y \in G$.

Although these definitions appear to be ad hoc, in fact the four properties of groups have been carefully chosen so that they will apply to many different aspects of mathematics. Here are some important examples of groups that appear on other contexts besides group theory:

Example 1.1
The set of integers $\mathbb{Z}$, with the binary operation being the sum of two numbers.

The identity element is 0, and $-x$ is the inverse of x. This forms an abelian group. □

Example 1.2
Consider the set of rational numbers, denoted by $\mathbb{Q}$. We will still use addition for our binary operation. This is also an abelian group. □

Example 1.3
Consider the set of all rational numbers except for 0. This time we will use multiplication instead of addition for our group operation. The identity element is now 1, and the inverse of an element is the reciprocal. This abelian group will be denoted by $\mathbb{Q}^*$. □

Example 1.4
Consider the set of all *linear* functions of the form $f(x) = mx + b$, with $m, b \in \mathbb{R}$, $m \neq 0$. (The $\mathbb{R}$ represents the real numbers.) We multiply two linear functions together by function composition. That is, if $f(x) = mx + b$ and $g(x) = nx + c$, then

$$f \cdot g = g(f(x)) = n(mx + b) + c = (mn)x + (nb + c).$$

Note that in $f \cdot g$, we do f first, then g, so that it appears reversed in $g(f(x))$. This group satisfies all of the group properties, but is not abelian. For example, if $f(x) = 2x + 3$ and $g(x) = 3x + 2$, then $f \cdot g = g(f(x)) = 6x + 11$, whereas $g \cdot f = f(g(x)) = 6x + 7$. □

DEFINITION 1.5 The number of elements in a group G is called the *order* of the group, and is denoted $|G|$. If G is has an infinite number of elements, we say that $|G| = \infty$.

Examples 1.1 though 1.4 have infinite order, and hence we cannot form multiplication tables for these groups. On the other hand, the *smallest* possible group is given by the following example.

Example 1.5
Consider the group containing just the identity element, $\{e\}$. We can have *Mathematica* give a multiplication table of this group by the following command:

InitGroup[e];
MultTable[{e}]

$\cdot$	e
e	e

We call this group the *trivial group*. The last *Mathematica* command introduces a new command—**InitGroup**[ ]. This command erases all previous groups, and designates the new identity element. The command

ClearDefs

erases all groups, and returns the dot to its standard definition. □

It takes a bit more work to define the trivial group in GAP so that e is the identity:

```
gap> f:=FreeGroup("e");;
gap> g:=f/[f.1];;
gap> e := g.1;;
gap> MultTable([e]);

*|e
-+-
e|e
```

The meaning behind the command `FreeGroup` will be dealt with in section 2.2.

Note that sometimes the operator $(\cdot)$ means addition, sometimes it means multiplication, and sometimes it means neither. Nonetheless, we can define x^n to mean x operated on itself n times. Thus,

$$x = x^1,$$

$$x \cdot x = x^2,$$

$$x \cdot x \cdot x = x^3,$$

etc.

We want to formally define x^n for any integer n. We let $x^0 = e$, the identity element. We then define, for $n > 0$,

$$x^n = x^{n-1} \cdot x.$$

By defining the nth power in terms of the previous power, we have defined x^n whenever n is a positive integer.

Finally, we can define negative powers by letting

$$x^{-n} = (x^n)^{-1} \qquad \text{if} \quad n > 0.$$

This is an *inductive* definition, since it defines each power in terms of a previous power. This type of definition works well for proving simple propositions about x^n.

PROPOSITION 1.6
If x is an element in a group G, and m and n are integers, then

$$x^{m+n} = x^m \cdot x^n.$$

PROOF If m or n are 0, this proposition is very easy to verify:

$$x^{m+0} = x^m = x^m \cdot e = x^m \cdot x^0, \qquad x^{0+n} = x^n = e \cdot x^n = x^0 \cdot x^n.$$

We will now prove the statement when m and n are positive integers. If n is 1, then we have

$$x^{m+1} = x^{(m+1)-1} \cdot x = x^m \cdot x^1,$$

using the inductive definition of the power of x.

We will now proceed by means of *induction.* That is, we will assume that the statement is true for $n = k - 1$, and then prove that it is then true for $n = k$. Then we will have that, since the statement is true for $n = 1$, and it is true for each number that follows, it must be true for all positive n.

Thus, we will assume that

$$x^{m+(k-1)} = x^m \cdot x^{k-1}.$$

But then

$$x^{m+k} = x^{m+k-1} \cdot x = x^m \cdot x^{k-1} \cdot x = x^m \cdot x^k.$$

Thus, by assuming the statement is true for $n = k - 1$, we found that it was also true for $n = k$. By induction, this proves that $x^{m+n} = x^m \cdot x^n$ for all positive n.

Once we have the statement true for positive m and n, we can take the inverse of both sides to give us

$$(x^{m+n})^{-1} = (x^n)^{-1} \cdot (x^m)^{-1}.$$

But by the definition of negative exponents, this is

$$x^{(-n)+(-m)} = x^{-n} \cdot x^{-m}$$

which, by letting $M = -n$ and $N = -m$, proves the proposition for the case of both exponents being negative.

Finally, if m and n have different signs, then $(m + n)$ will either have the same sign as $-n$, or the same sign as $-m$. If $(m + n)$ has the same sign as $-n$, then we have already shown that

$$x^m = x^{(m+n)+(-n)} = x^{m+n} \cdot x^{-n}.$$

So we have $x^m \cdot (x^{-n})^{-1} = x^{m+n} \cdot x^{-n} \cdot (x^{-n})^{-1}$, and hence $x^{m+n} = x^m \cdot x^n$.

If $(m + n)$ has the same sign as $-m$, then we have already shown that

$$x^n = x^{(-m)+(m+n)} = x^{-m} \cdot x^{m+n}.$$

So we have $(x^{-m})^{-1} \cdot x^n = (x^{-m}) \cdot x^{-m} \cdot x^{m+n}$, and hence $x^{m+n} = x^m \cdot x^n$.

Thus we have proven the proposition for all integers m and n. □

This last proof introduces an important method of proving theorems called *induction.* Induction is based on the simple fact that if a set of positive integers contains the number 1, and has the property that k is in the set whenever $k-1$ is, then the set must be all positive integers.

It is not hard to see why this must be true. If there were some positive integer not in the set, then there must be a *smallest* positive integer k that is not in the set. Since 1 is in the set, we see that $k > 1$, and since k is the smallest number not in the set, $k-1$ must be in the set. But the property of the set is that if $k-1$ is in the set, then k also is. So we have a contradiction, so there is no such k, meaning the the set is indeed all positive integers.

This gives us a powerful tool for proofs. In fact, we really do not need to introduce the variable k. To prove a statement for all positive integers n, we can first prove the statement is true for $n = 1$, and then we can assume that the statement is true for the previous case $n-1$. This extra information often gives us the leverage we need to be able to prove the statement is true for n. Here is another example of the use of induction.

PROPOSITION 1.7

If x is an element in a group G, and m and n are in $\mathbb{Z}$, then

$$(x^m)^n = x^{(mn)}.$$

PROOF Notice that this statement is trivial if $n = 0$ and $n = 1$:

$$(x^m)^0 = e = x^{m \cdot 0}, \qquad (x^m)^1 = x^m = x^{(m \cdot 1)}.$$

We will again proceed by means of induction, which means we can assume that the statement is true for the previous case, with n replaced by $n-1$. That is, we can assume that

$$(x^m)^{n-1} = x^{m \cdot (n-1)}.$$

Note that

$$(x^m)^n = (x^m)^{n-1} \cdot x^m = x^{m \cdot (n-1)} \cdot x^m$$

By proposition 1.6, this is equal to $x^{m \cdot (n-1)+m} = x^{mn}$.

So by induction, the proposition holds for positive n. To see that it holds for negative n as well, simply note that

$$(x^m)^n = ((x^m)^{-n})^{-1} = (x^{-mn})^{-1} = x^{mn}.$$

If n is negative, then $-n$ is positive, so the second step is valid. □

The principle of induction can easily be generalized. In proving a statement is true for the case n, not only can we assume that it is true for $n-1$, but also we can assume that the statement is true for all values smaller than n as well. Also, there is no reason why we must start with the number 1. Any other integer can be used for the starting point. That is, we first prove the statement is true for the case c. If we can then prove that the statement is true for n, utilizing the assumption that the statement is true for all numbers between c and $n-1$, then we have successfully proven that the statement is true for all integers greater than c. Problems 1.27 through 1.33 give some practice for using the principle of induction.

Problems for Chapter 1

Interactive Problems

1.1 If Terry was only allowed to do the dance steps **FlipRt** or **FlipLft**, could it get itself into all six possible positions? If possible, express the other four dance steps in terms of these two. Either the *Mathematica* command

InitTerry

or the GAP command

```
gap> InitTerry();
```

reloads Terry's group.

1.2 Repeat problem 1.1, only allow Terry to do only the steps **RotRt** and **RotLft**.

1.3 We saw that there were exactly four numbers less than 10 which were invertible modulo 10. For what other values of n are there exactly four numbers less than n which are invertible modulo n? Use *Mathematica*'s circle graph to graph the inverse functions.

1.4 According to the Chinese Remainder Theorem (1.3), there is a number less than 77 that is congruent to 4 Mod 11, and congruent to 6 Mod 7. Find this number, using either GAP or *Mathematica* to help.

1.5 The following *Mathematica* command creates a multiplication table of the five elements $\{e, a, b, c, d\}$. First execute this command:

InitGroup[e];
Define[a.a, e]; Define[a.b, c]; Define[a.c, d]; Define[a.d, b];
Define[b.a, d]; Define[b.b, e]; Define[b.c, a]; Define[b.d, c];

Define[c.a, b]; Define[c.b, d]; Define[c.c, e]; Define[c.d, a]; Define[d.a, c]; Define[d.b, a]; Define[d.c, b]; Define[d.d, e]; MultTable[{e, a, b, c, d}]

Notice that this multiplication table satisfies the "Latin square" property, hence this multiplication satisfies proposition 1.3. Does this set form a group? Why or why not?

Non-Interactive Problems

1.6 Suppose that Terry the Triangle has a friend who is a square. (Most of us have had such a friend from time to time.) How many dance steps would the square have? Construct a multiplication table of all of the square's dance steps. This group is referred to as D_4.

1.7 Suppose that Terry has a friend who is a regular tetrahedron. (A tetrahedron is a triangular pyramid.) How many dance steps would this tetrahedron have?

1.8 Using only the four basic properties of groups, prove that there can be only one identity element. That is, there cannot be two elements e and e' for which $x \cdot e = e \cdot x = x$ and $x \cdot e' = e' \cdot x = x$ for all $x \in G$.

1.9 Using only the four basic properties of groups, prove that an element cannot have two different inverses. That is, show that there cannot be to elements y and y' such that both $x \cdot y = e$ and $x \cdot y' = e$.

1.10 Prove that if a and b are two of Terry's dance steps, then there is a unique dance step x such that

$$x \cdot a = b.$$

This shows that every column in the multiplication table contains one and only one of each element.

1.11 If two of Terry's dance steps are chosen at random, what are the chances that these two dance steps will commute?

Hint: There are 36 ways of choosing two dance steps. Count the number of combinations that satisfy the equation $x \cdot y = y \cdot x$.

For problems **1.12** through **1.15**: Construct a multiplication table for the set of numbers modulo n.

Hint: Since these are the numbers that have multiplicative inverses modulo n, proposition 1.4 shows that the multiplication table has the same properties as Terry's dance steps.

1.12 $\{1, 2, 4, 5, 7, 8\}$, $n = 9$

1.13 $\{1, 3, 5, 9, 11, 13\}$, $n = 14$

1.14 $\{1, 5, 7, 11, 13, 17\}$, $n = 18$

1.15 $\{1, 5, 7, 11, 13, 17, 19, 23\}$, $n = 24$

1.16 Find the GCD of the numbers 24 and 42. Find two integers u and v such that $24u + 42v = \mathrm{GCD}(24, 42)$.

1.17 Find the GCD of the numbers 100 and 36. Find two integers u and v such that

$$100u + 36v = \mathrm{GCD}(100, 36).$$

Hint: Examine the multiples of 36, in particular the last two digits.

1.18 Find a positive integer $k < 35$ such that

$$k \equiv 1 \pmod 5 \quad \text{and} \quad k \equiv 0 \pmod 7.$$

Then find an integer $p < 35$ such that

$$p \equiv 0 \pmod 5 \quad \text{and} \quad p \equiv 1 \pmod 7.$$

Show how you can use p and k to compute a number n such that

$$n \equiv x \pmod 5 \quad \text{and} \quad n \equiv y \pmod 7$$

for a given x and y. The number n does not have to be less than 35.

1.19 Let u, v, and w be three positive integers that are *mutually* coprime. That is, each is coprime to the other two. Given any x, y, and z in $\mathbb{Z}$, prove that there is a unique number k such that

$$0 \leq k < u \cdot v \cdot w,$$

$$k \equiv x \pmod u,$$

$$k \equiv y \pmod v,$$

and

$$k \equiv z \pmod w.$$

Hint: Use the Chinese remainder theorem (1.3).

1.20 Suppose that S is a finite set (not necessarily a group) which is closed under the operator $(\cdot)$. Suppose also that the equation

$$a \cdot x = a \cdot y$$

holds if, and only if, $x = y$. Prove proposition 1.3 holds for the set S, even if S is not a group.

Hint: Use the pigeonhole principle.

1.21 Consider the set of all non-negative integers, with addition as the binary operation. Is this a group? Why or why not?

1.22 If G is a group such that $x^2 = e$ for all elements x in G, prove that G is commutative.

1.23 Let G be a group. Show that G is commutative if, and only if, $(a \cdot b)^2 = a^2 \cdot b^2$ for all a and b in G.

1.24 Let G be a finite group that contains an even number of elements. Show that there is at least one element besides the identity such that $a^2 = e$.

Hint: Show that there are an even number of elements for which $a^2 \neq e$.

1.25 Let G be a finite group. Show that there are an odd number of elements that satisfy the equation $a^3 = e$.

1.26 The following is a partially filled-in multiplication table for a group of order 8.

	a	b	c	d	e	f	g	h
a	b		d					c
b	g	e		h				
c						e	d	g
d		h		b			f	
e			c					
f			e			b		a
g	e	a			g		b	
h			a				c	

Fill in the remaining spaces in this multiplication table so that the resulting set forms a group.

Hint: Once the row and column of the identity element are filled in, the remaining table can be finished using only the Latin square property.

1.27 Use induction to prove that for all positive integers n,

$$1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2}.$$

1.28 Use induction to prove that for all positive integers n,

$$1 + 3 + 5 + \cdots + (2n - 1) = n^2.$$

1.29 Use induction to prove that for all positive integers n,

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

1.30 Use induction to prove that for all positive integers n,

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{n^2(n+1)^2}{4}.$$

1.31 Use induction to prove that for all positive integers n,

$$1 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + \cdots + n(n+1) = \frac{n(n+1)(n+2)}{3}.$$

1.32 Use induction to prove that for all positive integers n,

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \cdots + \frac{1}{n(n+1)} = \frac{n}{n+1}.$$

1.33 Use generalized induction to prove that all integers greater than 1 are either prime, or can be written as a product of primes.

This page intentionally left blank

Chapter 2

The Structure within a Group

2.1 Generators of Groups

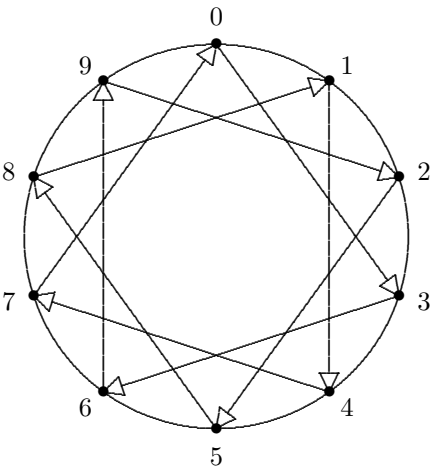
In this section we study *finite* groups, such as Terry's group, Z_n, and Z_n^*. By observing the properties of a single element within such a group, we gain insight on how to program *Mathematica*® or GAP to work with finite groups.

We begin with the group Z_{10}, which is loaded into *Mathematica* with the command

DefSumMod[10]

We can map each element x to the element $x \cdot 3$ with a circle graph

CircleGraph[{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, Add[3]]



This graph allows us to visualize powers of 3 in the group Z_{10}. If we follow the arrows starting with 0, we have the sequence $\{0, 3, 6, 9, 2, 5, 8, 1, 4, 7, 0 \ldots\}$. This tells us that

$$3^0 = 0, \qquad 3^1 = 3, \qquad 3^2 = 6, \qquad 3^3 = 9, \qquad 3^4 = 2, \quad \text{etc.}$$

Recall that for this group the dot represents addition, so an exponent would represent repeated addition. Note that every element in the group can be

expressed as a power of 3. This property does not hold for all elements, since the powers of 4 are seen to be $\{0, 4, 8, 2, 6, 0, 4, 8, \ldots\}$, which does *not* include all of the elements.

DEFINITION 2.1 We'll say that the element $g \in G$ is a *generator* of the group G if every element of G can be expressed as a power of g.

The natural question that arises is whether a given element is a generator of a group. This is not difficult for the group Z_n.

PROPOSITION 2.1
The generators of Z_n are precisely the integers between 0 and n that are coprime to n.

PROOF Suppose that g is a generator of Z_n. Then 1 is able to be expressed as a power of g, so we have that

$$g^v \equiv 1 \text{ (Mod } n)$$

for some v. Since the group action of Z_n is addition, raising to a power is equivalent to repeated addition, or standard multiplication. Thus, we have that

$$gv \equiv 1 \text{ (Mod } n).$$

By proposition 1.5, there is such a v if, and only if, g is coprime to n.

Now suppose that g is coprime to n. By proposition 1.5, there is a v such that

$$g^v = gv \equiv 1 \text{ (Mod } n).$$

So 1 can be expressed as a power of g. But 1 is a generator of Z_n, and so every element of Z_n can be expressed as a power of 1, say 1^w. Then that element can be written as $g^{(vw)} = (g^v)^w = 1^w$. So every element can be expressed as a power of g, hence g is a generator of Z_n. □

The count of numbers less than n that are coprime to n is called the *Euler totient function* of n, and is denoted $\phi(n)$. Thus, the number of generators of Z_n is precisely $\phi(n)$. A small table of this function up to $n = 36$ is given in table 2.1.

For larger values of n, we can use the *Mathematica* command **EulerPhi** or the GAP command `Phi`.

EulerPhi[60]

```
gap> Phi(60);
16
```

TABLE 2.1: Table of $\phi(n)$

n	$\phi(n)$	n	$\phi(n)$	n	$\phi(n)$	n	$\phi(n)$
1	1	10	4	19	18	28	12
2	1	11	10	20	8	29	28
3	2	12	4	21	12	30	8
4	2	13	12	22	10	31	30
5	4	14	6	23	22	32	16
6	2	15	8	24	8	33	20
7	6	16	8	25	20	34	16
8	4	17	16	26	12	35	24
9	6	18	6	27	18	36	12

Hence, there are 16 generators of Z_{60}. Both programs use the following formula for the totient function based on the prime factorization of the number.

THEOREM 2.1: The Totient Function Theorem
If the prime factorization of n is given by

$$n = p_1^{r_1} \cdot p_2^{r_2} \cdots p_k^{r_k},$$

where $p_1, p_2, p_3, \ldots, p_k$ are distinct primes, and $r_1, r_2, r_3, \ldots, r_k$ are positive integers, then the count of numbers less then n which are coprime to n is

$$\phi(n) = (p_1 - 1) \cdot p_1^{(r_1-1)} \cdot (p_2 - 1) \cdot p_2^{(r_2-1)} \cdot \cdots \cdot (p_k - 1) \cdot p_k^{(r_k-1)}.$$

PROOF To begin, let us show that if p is a prime, then $\phi(p^r) = (p-1)p^{r-1}$.

Note that the only numbers that are *not* coprime to p^r will be multiples of p. So of the numbers between 1 and p^r, exactly $1/p$ of them will be multiples of p. The remaining $(1 - 1/p) \cdot p^r$ will be coprime, and this can be simplified to $(p-1)p^{r-1}$.

Next we want to show that if n and m are coprime, then $\phi(nm) = \phi(n)\phi(m)$. Let A denote the set of numbers that are less than n, but coprime to n. Let B denote the set of numbers that are less than m, but coprime to m.

Then for any number coprime to $n \cdot m$, that number, modulo n, must be in the set A, while that number, modulo m, must be in B. Yet for every a in A and b in B, there is, by the Chinese remainder theorem, a unique number less than $n \cdot m$ that is equivalent to a (Mod n) and b (Mod m). This number will be coprime to both n and m, and hence will be coprime to $n \cdot m$.

Therefore, we have a one-to-one correspondence between ordered pairs (a, b), where a is in A, and b is in B, and numbers coprime to $n \cdot m$. Thus, we have

$$\phi(n \cdot m) = \phi(n) \cdot \phi(m).$$

Finally, we can combine these results together. By simply noting that if

$$n = p_1^{r_1} \cdot p_2^{r_2} \cdots p_k^{r_k},$$

then $p_1^{r_1}$, $p_2^{r_2}$, $p_3^{r_3}, \ldots, p_k^{r_k}$ will all be coprime. Hence, we can find ϕ for each of these terms, and multiply them together, giving us our formula. □

We can also consider finding generators for the groups of the form Z_n^*. For example Z_{10}^* has four elements, $\{1, 3, 7, 9\}$, and we find that two of these are generators, 3 and 7. But Z_8^* also has four elements, $\{1, 3, 5, 7\}$, yet none of these elements are generators of the group! This becomes apparent as we look at the multiplication table for Z_8^*.

```
gap> MultTable([1,3,5,7]);
```

or, in *Mathematica*,

DefMultMod[8]
MultTable[{1, 3, 5, 7}]

·	1	3	5	7
1	1	3	5	7
3	3	1	7	5
5	5	7	1	3
7	7	5	3	1

Notice that the square of every element is equal to 1. Hence no element of Z_8^* can generate the whole group.

DEFINITION 2.2 We say a group is *cyclic* if there is one element that can generate the entire group.

Although we have seen an example of a finite group that is not cyclic, we will later see that the structure of *any* finite abelian group can be expressed in terms of the cyclic groups.

Even when a group is not cyclic, we sometimes can find *two* elements by which every element of the group can be expressed. For example, consider the two elements 3 and 5 from the group Z_8^*. Since $1 = 3 \cdot 3$ and $7 = 3 \cdot 5$, we find that all four elements of the group can be written as some *combination* of 3 and 5. We say that the *set* $\{3, 5\}$ generates the group.

Finally, consider the group of the dancing triangle, whose multiplication table is given in table 1.2. By experimenting, we find that no single element can generate the entire group. However, there are many ways in which we can have *two* elements generating the entire group. For example, if we pick the two elements **RotRt** and **Spin**, we find that the other four elements can be expressed in terms of these two: **Stay** = **Spin**·**Spin**, **FlipRt** = **Spin**·**RotRt** **FlipLft** = **RotRt** · **Spin**, and **RotLft** = **RotRt** · **RotRt**.

One of the keys for entering a group into *Mathematica* is finding one or two elements (or sometimes even three are needed) that will generate the entire group. This information begins to reveal the structure of the group itself.

2.2 Defining Finite Groups in *Mathematica* and GAP

For some groups there is a single element that generates the entire group, whereas in other groups two or more elements are required. In this section we will show how a finite group can be entered into *Mathematica* or GAP using a set of elements that generates the group. We will begin with a cyclic group Z_n which has a single generator which we will call x. By the pigeonhole principle, the sequence of n elements

$$\begin{aligned} e &= x^0, \\ x &= x^1, \\ x \cdot x &= x^2, \\ x \cdot x \cdot x &= x^3, \\ \cdots &\quad \cdots \\ x \cdot x \cdot x \cdot \cdots \cdot x &= x^{(n-1)}, \end{aligned}$$

must mention every element of Z_n exactly once. This gives us a way to label the elements of Z_n in terms of the generator x. We also find that $x^n = e$. Thus, we can define the group Z_n merely by saying "x is a generator of the group, and n is the lowest number such that x^n is the identity."

There are *Mathematica* routines that allow us to quickly make these definitions. The two statements

```
InitGroup[e]
Define[x^5, e]
```

define x^5 to be the identity e. This alone is sufficient to define the group Z_5. To view this group, we use the command

```
Z5 = Group[{x}]
```

which gives a list of all of the elements in the group, and assigns this list to the identifier **Z5**. The multiplication table for this group produced by the **MultTable** command is shown in table 2.2.

Once the group is defined, we can multiply elements of the group with the dot, and *Mathematica* will simplify them.

```
x^4 . x^4
x . x . x
```

Notice that the elements can be entered into *Mathematica* using the power notation, but they are displayed as a repeated product. Although the notation $\{0, 1, 2, 3, 4\}$ is more concise for this particular example, the use of generators is more versatile, since almost all finite groups can be expressed in an easy way using generators.

To define the same group in GAP using generators, we begin by defining

TABLE 2.2: Table of Z_5

$\cdot$	e	x	$x \cdot x$	$x \cdot x \cdot x$	$x \cdot x \cdot x \cdot x$
e	e	x	$x \cdot x$	$x \cdot x \cdot x$	$x \cdot x \cdot x \cdot x$
x	x	$x \cdot x$	$x \cdot x \cdot x$	$x \cdot x \cdot x \cdot x$	e
$x \cdot x$	$x \cdot x$	$x \cdot x \cdot x$	$x \cdot x \cdot x \cdot x$	e	x
$x \cdot x \cdot x$	$x \cdot x \cdot x$	$x \cdot x \cdot x \cdot x$	e	x	$x \cdot x$
$x \cdot x \cdot x \cdot x$	$x \cdot x \cdot x \cdot x$	e	x	$x \cdot x$	$x \cdot x \cdot x$

```
gap> f := FreeGroup("x");
<free group on the generators [ x ]>
gap> AssignGeneratorVariables(f);
#I  Assigned the global variables [ x ]
```

There may also be a warning message saying that x was defined to be something else, but just ignore this warning. This defines a group with a generator of x, and in fact inverses are automatically defined. GAP uses a star instead of a dot for multiplication.

```
gap> x^5 * x^-7;
x^-2
```

At this point, though, we have an infinite group. We still need to declare that $x^5 = e$. We do this by defining another group g as follows:

```
gap> g := f/[x^5];
<fp group on the generators [x]>
gap> AssignGeneratorVariables(g);
#I  Global Variable 'x' already defined and will be overwritten
#I  Assigned the global variables [ x ]
```

The `f/[x^5]` is GAP's way of declaring x^5 to be the identity. At this point, g is defined to be the new group. To find its size and list its elements, we can use the commands

```
gap> Size(g);
5
gap> List(g);
[ <identity ...>, x, x^2, x^3, x^4 ]
```

The identity element of the group is listed as `<identity ...>`, which of course is the yet unnamed identity element. But the multiplication table can still be displayed.

```
gap> MultTable(g);
```

```
*  |e    x    x^2  x^3  x^4
---+----------------------
e  |e    x    x^2  x^3  x^4
x  |x    x^2  x^3  x^4  e
x^2|x^2  x^3  x^4  e    x
x^3|x^3  x^4  e    x    x^2
x^4|x^4  e    x    x^2  x^3
```

When the table is displayed, the identity element is displayed as e, making the table more concise that its *Mathematica* counterpart. The identity element can be given any name by changing the variable `DisplayIdentity`, which has a default setting of `"e"`. If we multiply elements together,

```
gap> x^4 * x^4;
x^8
```

we find it doesn't simplify yet. If we give the command

```
gap> SetReducedMultiplication(g);
gap> x^4 * x^4;
x^-2
```

then GAP will simplify products, but not always to the same product that *Mathematica* will simplify it to.

For an example requiring two generators, consider Z_8^*, which can be generated by $a = 3$ and $b = 5$. This group can be entered into *Mathematica* with the commands:

InitGroup[**e**]
Define[**a.a, e**]
Define[**b.b, e**]
Define[**b.a, a.b**]
G = Group[{**a, b**}]

Note that we needed an extra **Define** statement to let *Mathematica* know that a and b commute with each other. We can actually define several groups at the same time in *Mathematica*, as long as we use the same symbol for the identity element. However, the command

InitGroup[**e**]

clears all previously defined groups. This group can be defined in GAP in a similar way.

```
gap> f:=FreeGroup("a","b");;
gap> AssignGeneratorVariables(f);
#I  Assigned the global variables [ a, b ]
gap> h:=f/[a^2,b^2,a*b*a*b];;
gap> List(h);
[ <identity ...>, a, b, a*b ]
gap> MultTable(h);

*  |e    a    b    a*b
---+-------------------
e  |e    a    b    a*b
a  |a    e    a*b  b
b  |b    a*b  e    a
a*b|a*b  b    a    e
```

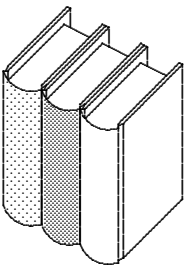


FIGURE 2.1: Three books that can be rearranged

By using a different identifier h for the new group, we still have the older group g defined in terms of the different generator x. Notice that in GAP, we "divide" f by a list of combinations that should reduce to the identity element. Instead of declaring $b \cdot a = a \cdot b$ as we did in *Mathematica*, we are declaring that $a \cdot b \cdot a \cdot b = e$. In problem 2.18, we see that this is equivalent to saying $b \cdot a = a \cdot b$.

To do group operations, we still need the following commands:

```
gap> AssignGeneratorVariables(h);
#I  Global variable 'a' already defined and will be overwritten
#I  Global variable 'b' already defined and will be overwritten
#I  Assigned the global variables [ a, b ]
gap> SetReducedMultiplication(h);
gap> b*a;
a^-1*b^-1
```

Clearly GAP's definition of simplest form is not the same as *Mathematica*! In GAP's "dictionary," a^{-1} comes before a, and `SetReducedMultiplication` will put the element into a form with the fewest multiplications, and for a tie-breaker, GAP finds the form that comes first in a dictionary. Since GAP considers a^{-1} coming before a, $a^{-1} * b^{-1}$ is before the simpler looking $a * b$.

Here is another example of a group. Suppose we have three different books on a shelf, and we consider rearrangements of the books. Such a group of arrangements can be entered in *Mathematica* with the command

InitBooks[3]

which begins by showing three differently colored books, as in figure 2.2. Two ways we could rearrange the books are to swap the first two books, or move the first book to the other end, sliding the other two books to the left. These two operations can be animated in *Mathematica* by

MoveBooks[First]
MoveBooks[Left]

By letting e be the identity element, a be the first rearrangement, and b be the rearrangement moving the books to the left, we find that all possible

permutations of the books are generated by a and b. Since $a^2 = b^3 = e$, and the combination $a \cdot b$ essentially switches the last two books, we see that $(a \cdot b)^2 = e$. Thus, we can define this group in GAP by

```
gap> f := FreeGroup("a","b");;
gap> a := f.1;;
gap> b := f.2;;
gap> g := f/[a^2, b^3, (a*b)^2];;
gap> a := g.1;;
gap> b := g.2;;
gap> List(g);
[ <identity ...>, a, b, a*b, a*b*a, b*a ]
gap> MultTable(g);

*     |e      a      b      a*b    a*b*a  b*a
------+-------------------------------------
e     |e      a      b      a*b    a*b*a  b*a
a     |a      e      a*b    b      b*a    a*b*a
b     |b      b*a    a*b*a  a      e      a*b
a*b   |a*b    a*b*a  b*a    e      a      b
a*b*a |a*b*a  a*b    e      b*a    b      a
b*a   |b*a    b      a      a*b*a  a*b    e
```

Notice that instead of using `AssignGeneratorVariables`, we set a to `f.1`, meaning the first generator of f, and set b to f's second generator, `f.2`. Later, we set a and b to the generators of g. This is precisely what the `AssignGeneratorVariables` command did.

To define this group in *Mathematica*, we begin with

InitGroup[e]
Define[a^2, e]
Define[b^3, e]

We also have to define $b \cdot a$ in terms of $a \cdot b$, just as we did in defining Z_8^*. We observe that $b \cdot a = a \cdot b \cdot b$ instead of $a \cdot b$. So to finish defining this group, we have

Define[b.a, a.b.b]
G = Group[{a, b}]

This group is called S_3, the permutation group on three objects. (Obviously it makes no difference what the three objects are. Books are just one possibility.) Table 2.3 shows the multiplication table.

Although many of the properties of groups can be verified by looking at the table, the associativity is not obvious. We can have *Mathematica* verify that the associative property holds for G with the command

CheckGroup[G]

If we try to take an inverse of an element using *Mathematica*,

(a.b)^(–1)
$b^{-1} \cdot a^{-1}$

TABLE 2.3: Multiplication table for S_3

$\cdot$	e	a	b	$a \cdot b$	$b \cdot b$	$a \cdot b \cdot b$
e	e	a	b	$a \cdot b$	$b \cdot b$	$a \cdot b \cdot b$
a	a	e	$a \cdot b$	b	$a \cdot b \cdot b$	$b \cdot b$
b	b	$a \cdot b \cdot b$	$b \cdot b$	a	e	$a \cdot b$
$a \cdot b$	$a \cdot b$	$b \cdot b$	$a \cdot b \cdot b$	e	a	b
$b \cdot b$	$b \cdot b$	$a \cdot b$	e	$a \cdot b \cdot b$	b	a
$a \cdot b \cdot b$	$a \cdot b \cdot b$	b	a	$b \cdot b$	$a \cdot b$	e

we find that *Mathematica* uses proposition 1.2 to express the answer in terms of a^{-1} and b^{-1}. But unlike GAP, *Mathematica* does not yet know the inverses of a and b. We can remedy the situation with two more **Define** commands:

Define[a^(–1), a]
Define[b^(–1), b.b]

Mathematica can now find the inverse of any element x by entering either x^{-1} or $1/x$.

Although the two programs display the elements of the group differently, we can get GAP to display a table very similar to *Mathematica*'s with the commands

```
gap> L := ListGroup(g);
[ <identity ...>, a, b, a*b, b^2, a*b^2 ]
gap> MultTable(L);
```

which will force the elements to be in a certain order in the table, and expressed in a certain way.

The multiplication tables for Terry's group and S_3 are very similar. By color coding the elements in the table, we see that the color patterns of the two multiplication tables are identical. Thus, these two groups behave in exactly the same way, even though the elements have different names. We say that these groups are *isomorphic.* We will cover isomorphic groups in chapter 4.

Finally, let us consider the group of rotations on the octahedron. *Mathematica*'s command

ShowOctahedron

displays a colored octahedron like the one in figure 2.2. There are eight triangles forming this solid. Three ways of rotating this figure are given by

RotateOctahedron[a]
RotateOctahedron[b]
RotateOctahedron[c]

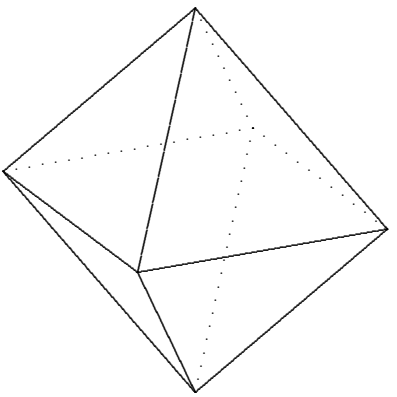


FIGURE 2.2: Octahedron with eight equilateral triangles

The first of these flips the front horizontal edge, turning it upside down. The second rotates the closest face counterclockwise, while the third rotates the closest vertex clockwise. If we let e be the identity element of this group, it is easy to see that

$$a^2 = e, \qquad b^3 = e, \qquad c^4 = e,$$

and hence

$$a^{-1} = a, \qquad b^{-1} = b^2, \qquad c^{-1} = c^3.$$

After some experimenting, we find that $b \cdot a \cdot b \cdot a = e$, $c \cdot b \cdot c \cdot c \cdot a = e$, and $c \cdot a \cdot c^3 \cdot a \cdot b = e$. From these identities, we can come up with the identities

$$b \cdot a = (b \cdot a)^{-1} = a^{-1} \cdot b^{-1} = a \cdot b^2.$$

$$c \cdot b = (c \cdot c \cdot a)^{-1} = a^{-1} \cdot c^{-1} \cdot c^{-1} = a \cdot c^3 \cdot c^3 = a \cdot c^2 \cdot c^4 = a \cdot c^2.$$

$$c \cdot a = (c^{-1} \cdot a \cdot b)^{-1} = b^{-1} \cdot a^{-1} \cdot c = b^2 \cdot a \cdot c = b \cdot a \cdot b^2 \cdot c = a \cdot b^4 \cdot c = a \cdot b \cdot c.$$

This allows us to define $b \cdot a$, $c \cdot a$, and $c \cdot b$ in terms of operations that are performed in *alphabetical order*. This is the key to defining a group in *Mathematica*.

```
InitGroup[e];
Define[a^2, e]
Define[b^3, e]
Define[c^4, e]
Define[1/a, a]
Define[1/b, b^2]
Define[1/c, c^3]
Define[b.a, a.b.b]
Define[c.a, a.b.c]
Define[c.b, a.c.c]
```

G = Group[{a, b, c}]

$\{e,\ a,\ b,\ c,\ a\cdot b,\ a\cdot c,\ b\cdot b,\ b\cdot c,\ c\cdot c,\ a\cdot b\cdot b,\ a\cdot b\cdot c,\ a\cdot c\cdot c,\ b\cdot b\cdot c,$
$b\cdot c\cdot c,\ c\cdot c\cdot c,\ a\cdot b\cdot b\cdot c,\ a\cdot b\cdot c\cdot c,\ a\cdot c\cdot c\cdot c,\ b\cdot b\cdot c\cdot c,\ b\cdot c\cdot c\cdot c,$
$a\cdot b\cdot b\cdot c\cdot c,\ a\cdot b\cdot c\cdot c\cdot c,\ b\cdot b\cdot c\cdot c\cdot c,\ a\cdot b\cdot b\cdot c\cdot c\cdot c\}$

By expressing the product of any two generators in terms of a combination in alphabetical order, *Mathematica* will make replacements in any combination until it is finally a combination of generators in alphabetical order, and then stop. We will cover the details of this process in section 8.3.

We call this group the *octahedral* group. The command

Length[G]

shows this group has 24 elements. This group is too large to print a complete multiplication table, but *Mathematica* is able to produce a color-coded table for groups of up to 27 elements.

The corresponding GAP commands for this group are

```
gap> f:=FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> g:=f/[a^2, b^3, c^4, b*a*b*a, c*b*c*c*a, c*a*c^3*a*b];;
gap> Size(g);
24
gap> a:=g.1;; b:=g.2;; c:=g.3;;
gap> SetReducedMultiplication(g);
gap> G := List(g);
[ <identity ...>, a^-1*b^-1, b, a^-1*b^-1*c, c^-2, a^-1, c,
  b^-1*c^-1*b^-1, b^-1, c*a^-1, c^-1*a^-1, a^-1*b^-1*c^-1,
  a^-1*b, b^-1*c, c*b, c^-1, a^-1*c, b^-1*c^-2, a^-1*b*c^-1,
  b*c, c^-1*b^-1, b^-1*c^-1, a^-1*c^-1, b*c^-1 ]
```

Unfortunately, the multiplication table for this group is much too big for the screen in GAP. However, we can still multiply two elements in the list

```
gap> G[4]*G[7];
b^-1*c^-1*b^-1
```

and see that we will always get another member of the list. This group will be an important example later on.

Notice that unlike *Mathematica*, we did *not* have to find $b\cdot a$, $c\cdot a$, and $c\cdot b$ in terms of elements multiplied in alphabetical order. This shows that entering groups in GAP is actually more flexible than with *Mathematica*, which uses a totally different algorithm.

2.3 Subgroups

A natural question to ask is whether we can have a smaller group inside of a particular group. We begin by saying that H is a *subset* of a group G,

denoted $H \subseteq G$, if H consists only of the elements of G. The empty set $\{\ \}$ is always considered to be a subset, but we will restrict our attention to non-empty subsets.

DEFINITION 2.3 We say that H is a *subgroup* of G if H is a non-empty subset of G and H is a group with respect to the operation $(\cdot)$ of G.

To see if H is a group, we must test all four of the group properties. But the associative property of H is guaranteed because the original group G is associative. The remaining three properties,

1. H is closed under multiplication. That is, $x \cdot y \in H$ whenever x and $y \in H$.
2. The identity element of G is in H.
3. Every element of H has its inverse in H. That is, $x^{-1} \in H$ whenever $x \in H$.

can be combined into one simple test.

PROPOSITION 2.2
Let $H \subseteq G$ and $H \neq \{\ \}$. Then H is a subgroup of G if, and only if, we have

$$x \cdot y^{-1} \in H \qquad \textit{for all} \qquad x, y \in H.$$

PROOF First of all, we need to see that if H is a subgroup, then $x \cdot y^{-1}$ is in H whenever x and y are in H. By property (3), y^{-1} is in H, and so by property (1), $x \cdot y^{-1}$ is in H.

Conversely, let us suppose that $H \subseteq G$, $H \neq \{\ \}$, and whenever $x, y \in H$, then $x \cdot y^{-1} \in H$. We need to see that properties (1) through (3) are satisfied.

Since H is not the empty set, there is an element x in H, and so $x \cdot x^{-1} = e$ is in H. Thus, property (2) holds.

Next, we have that if y is in H, then $e \cdot y^{-1} = y^{-1}$ is in H, and so property (3) holds.

Finally, if x and y are in H, then y^{-1} is in H, and so $x \cdot (y^{-1})^{-1} = x \cdot y$ is in H. Thus, property (1) also holds. □

Let us look at S_3, defined in *Mathematica* by the commands

```
InitGroup[e];
Define[a^2, e]
Define[b^3, e]
Define[b.a, a.b.b]
Define[1/a, a]
```

Define[1/b, b^2]
G = Group[{a, b}]

or by the GAP commands

```
gap> f:=FreeGroup("a","b");; a:=f.1;; b:=f.2;;
gap> g:=f/[a^2, b^3, (a*b)^2];; a:=g.1;; b:= g.2;;
gap> G := ListGroup(g);
[ <identity ...>, a, b, a*b, b^2, a*b^2 ]
```

We can find smaller groups within this one, such as

$$H = \{e, b, b^2\}.$$

It is easy to see that if x and y are in H, then $x \cdot y^{-1}$ is in H. Therefore, this is a subgroup.

Next, consider the group $\mathbb{Z}$. If we let k be any integer then we can let

$$k\mathbb{Z} = \{k \cdot x \quad | \quad x \in \mathbb{Z}\}$$

denote the multiples of k. Since the difference of two multiples of k is again a multiple of k, $k\mathbb{Z}$ is a subgroup of $\mathbb{Z}$.

If we take the intersection $H \cap K$ of two subgroups of G, we can ask whether we will obtain another subgroup of G. For both *Mathematica* and GAP, this is done by the command `Intersection`. For example, we can take the intersection of two sets

H = {e, b, b^2}
K = {e, a}
Intersection[H, K]

or in GAP

```
gap> e := Identity(g);
<identity ...>
gap> H := [e, b, b^2];;
gap> K := [e, a];;
gap> Intersection(H, K);
[ <identity ...> ]
```

to find the set of all elements in common with H and K. Note that sets are entered using curly braces in *Mathematica*, but with square brackets in GAP. Moreover, we can consider taking the intersection of a collection of many sets. If we let

```
gap> L := [[e, a, b], [e, a*b, b], [e, a, b, b^2]];;
```

L = {{e, a, b}, {e, a·b, b}, {e, a, b, b^2}}

then L represents a "set of sets." We can take the intersection of all of the sets in this collection with the command

Intersection[L]

or

```
gap> Intersection(L);
[ <identity ...>, b ]
```

The mathematical notation for this intersection is

$$\bigcap_{H \in L} H.$$

PROPOSITION 2.3
Given a group G and a non-empty collection of subgroups, donated by L, then the intersection of all of the subgroups in the collection

$$H^* = \bigcap_{H \in L} H$$

is a subgroup of G.

PROOF First of all, note that H^* is not the empty set, since the identity element is in each H in the collection. We now can apply proposition 2.2. Let x and y be two elements in H^*. Then, for every $H \in L$ we have $x, y \in H$. Since each H is a subgroup of G, we have

$$x \cdot y^{-1} \in H.$$

Therefore, $x \cdot y^{-1}$ is in H^*, and so H^* is a subgroup of G. □

This proposition allows us to generate a subgroup of G from any subset of G.

DEFINITION 2.4 Given a subset S of a group G, we define the *subgroup generated by S* to be

$$[S] = \bigcap_{H \in L} H$$

where L denotes the collection of subgroups of G that contain the set S.

Actually, $[S]$ is the *smallest* subgroup of G that contains S. Hence, we can determine $[S]$ another way. It is clear that $[S]$ contains all of the products of the form

$$x_1 \cdot x_2 \cdot x_3 \cdot \cdots \cdot x_n,$$

where either

$$x_k \in S \quad \text{or} \quad x_k^{-1} \in S \qquad (1 \leq k \leq n).$$

But the set of all such products forms a subgroup H of G that contains S. Thus, $H = [S]$.

The command `Group` finds $[S]$ for any set S. Thus, we can find the subgroup of S_3 generated by the element b by the *Mathematica* command

Group[{b}]

which produces the subgroup $\{e, b, b^2\}$ we observed before. The corresponding GAP commands are

```
gap> Group(b);
Group([ b ])
gap> List(last);
[ <identity ...>, b, b^2 ]
```

Notice that the `Group` command in GAP did not automatically list out the elements in the subgroup. We needed an extra `List` command to see the elements. The subgroup generated by the set $\{b, a \cdot b\}$ is

```
gap> List(Group(b, a*b));
[ <identity ...>, b*a*b, b, a*b, b^2, b^2*a*b ]
```

or

Group[{b, a.b}]

which produces the entire group. Note that if `SetReducedMultiplication` is not entered in GAP, the elements may appear in nonstandard combinations. Had we entered

```
gap> SetReducedMultiplication(g);
gap> List(Group(b, a*b));
[ <identity ...>, a^-1, b, a^-1*b, b^-1, a^-1*b^-1 ]
```

we would get exactly the same thing as `List(g)`.

Let's look at a larger group. The following *Mathematica* and GAP commands reload the octahedral group of order 24:

InitGroup[e];
Define[a^2, e]; Define[b^3, e]; Define[c^4, e]
Define[1/a, a]; Define[1/b, b^2]; Define[1/c, c^3]
Define[b.a, a.b.b]; Define[c.a, a.b.c]; Define[c.b, a.c.c]
G = Group[{a, b, c}]

```
gap> f:=FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> g:=f/[a^2, b^3, c^4, b*a*b*a, c*b*c*c*a, c*a*c^3*a*b];;
gap> a:=g.1;; b:=g.2;; c:=g.3;;
gap> SetReducedMultiplication(g);
gap> h := List(g);
[ <identity ...>, a^-1*b^-1, b, a^-1*b^-1*c, c^-2, a^-1, c,
  b^-1*c^-1*b^-1, b^-1, c*a^-1, c^-1*a^-1, a^-1*b^-1*c^-1,
  a^-1*b, b^-1*c, c*b, c^-1, a^-1*c, b^-1*c^-2, a^-1*b*c^-1,
  b*c, c^-1*b^-1, b^-1*c^-1, a^-1*c^-1, b*c^-1 ]
```

With the command

Group[{**c**}]

or

```
gap> List(Group(c));
[ <identity ...>, c^-2, c, c^-1 ]
```

we find that $[\{c\}]$ is a subgroup of order 4. Likewise, $[\{b, c\}]$ produces the whole group. Hence, the octahedral group can be generated in GAP with just two of the elements. For convenience, we originally used three elements to define the group in *Mathematica*. Besides, it is easier to put the octahedron back into its original position using three types of rotations instead of just two.

Finally, the subgroup

Group[{**a, b**}]

or

```
gap> List(Group(a,b));
[ <identity ...>, a^-1*b^-1, b, a^-1, b^-1, a^-1*b ]
```

is simply another copy of the group S_3. Thus, there is a copy of S_3 inside of the octahedral group. Notice that in GAP, the set of elements does not have to be enclosed as a set (this is optional), whereas *Mathematica* does require the elements to be in a set, even if there is only one element.

Let us now consider the cyclic subgroups of a group G. Notice that if we pick any element x of G, then $[\{x\}]$ will always be a cyclic subgroup of G. This subgroup is usually denoted by $[x]$.

DEFINITION 2.5 Let G be a group and let x be an element in G. We define the *order* of x to be $|[x]|$. That is, if $[x]$ is finite the order of x is the number of elements in $[x]$. If $[x]$ is an infinite group we define the order of x to be infinity.

PROPOSITION 2.4
Suppose that the element x has finite order n. Then n is the smallest positive integer such that $x^n = e$. Furthermore,

$$[x] = \{e, x, x^2, x^3, \ldots, x^{n-1}\}.$$

PROOF Since $[x]$ is finite, not all of the elements $\{x^0, x^1, x^2, x^3, x^4, \ldots\}$ can be distinct. Suppose that $x^p = x^q$ for two integers p and q, with $p > q$. Then $x^{(p-q)} = e$ and $(p-q) > 0$. So there exists a positive integer r such that $x^r = e$. We can let n be the smallest such integer. We want to prove that

$$[x] = \{e = x^0, x^1, x^2, x^3, \ldots, x^{n-1}\}$$

with these elements distinct. Indeed, if $x^p = x^q$ with $0 \leq q < p \leq n-1$, then $x^{p-q} = e$ and $0 < p - q < n$, which contradicts the definition of n. Therefore, the elements in

$$\{e = x^0, x^1, x^2, x^3, \ldots, x^{n-1}\}$$

are all distinct.

Finally, we need to show that if y is in $[x]$, then there exists a q such that $x^q = y$, with $0 \leq q \leq n-1$. But $y = x^k$ for some $k \in \mathbb{Z}$. We can define $q = k \pmod{n}$. Then $0 \leq q \leq n-1$ and furthermore, there is an integer r such that $k - q = n \cdot r$. Thus,

$$y = x^k = x^{(nr+q)} = (x^n)^r \cdot x^q = e^r \cdot x^q = x^q.$$

So every element of $[x]$ is of the form x^q, with $0 \leq q \leq n-1$. □

PROPOSITION 2.5

Suppose that x has infinite order. Then x^n is not the identity element for all nonzero integers n. Furthermore,

$$[x] = \{\ldots, x^{-3}, x^{-2}, x^{-1}, x^0 = e, x^1, x^2, x^3, \ldots\},$$

where the powers of x are all distinct.

PROOF Suppose that $x^n = e$ for some nonzero n. It suffices to consider the case $n > 0$, for if $x^n = e$, then $x^{-n} = e$.

By exactly the same reasoning as was used to prove proposition 2.4, we see that

$$[x] = \{e = x^0, x^1, x^2, x^3, \ldots, x^{n-1}\}.$$

But this contradicts the fact that $[x]$ was infinite. Therefore, $x^n = e$ only if $n = 0$.

Moreover, if $x^p = x^q$, then $x^{p-q} = e$ and so $p - q = 0$ by what we have just proved. Thus, the powers of x are all distinct. □

Even though the group in proposition 2.5 cannot be defined in *Mathematica* because it is infinite, it *can* be defined in GAP. In fact, we defined an infinite group in the process of defining all of the other groups. If we have x as the generator of an infinite group, then the group is defined by the following:

```
gap> f:=FreeGroup("x");; x := f.1;;
gap> Size(f);
infinity
gap> x^4 * x^-7;
x^-3
```

Granted, we cannot display all of the elements as we did for the other groups (`List(f)` produces an error message), but we can still multiply elements of this group.

Because of propositions 2.4 and 2.5, we know that any cyclic group G is either a finite group

$$G = \{e, x, x^2, x^3, \ldots, x^{n-1}\}$$

which resembles the group Z_n, or is an infinite group

$$G = \{\ldots, x^{-3}, x^{-2}, x^{-1}, x^0 = e, x^1, x^2, x^3, \ldots\},$$

which resembles the group $\mathbb{Z}$.

We can use *Mathematica* or GAP to quickly find the order of any element in the group. For example, to find the order of the element $b \cdot c$ of the octahedral group (which was not erased by the infinite group, since it used different generators), we type

Length[Group[{b.c}]]

or

```
gap> Order(b*c);
4
```

to see that the order of this element is 4. We can also use *Mathematica* to find the number of elements of a group of a given order. For example, we can find the number of elements of order 2 by squaring all of the elements, and counting the number of times the identity appears. Of course the identity squared will be the identity, which we do not count. For example, the number of elements of order 2 of the group Z_{12}

DefSumMod[12]
G = Group[{1}]

can by found by the command

G^2
{0, 2, 4, 6, 8, 10, 0, 2, 4, 6, 8, 10}

which computes the square of each element in the group. Only 0 and 6 satisfy $x^2 = 0$, and 0 is of order 1. Thus, there is only one element of order 2 in the group Z_{12}.

This trick of raising the entire list to a power will not work in GAP. However, we can apply a function to all of the elements in a list with a special feature of GAP's `List` command.

```
gap> f:=FreeGroup("x");; x := f.1;;
gap> g:=f/[x^12];; x:=g.1;;
gap> Z12 := List(g);
[ <identity ...>, x^9, x^4, x^6, x, x^3, x^8, x^10, x^5, x^7,
  x^2, x^11 ]
gap> List(Z12, Order);
[ 1, 4, 3, 2, 12, 4, 3, 6, 12, 12, 6, 12 ]
```

When the `List` command has a second argument, it applies this function to every element in the list. This is a handy shortcut for sending each element of the list into any function. Using either *Mathematica* or GAP, we see that there is only one element of order 2, two elements each of order 3, 4, and 6, and four elements of order 12.

It is apparent that finding the number of elements of order k involves finding the number of solutions to the equation $x^k = e$. To help us find the number of solutions for a cyclic group, let us first prove the following proposition about modular multiplication.

PROPOSITION 2.6
Let n and k be two positive integers. Then

$$x \cdot k \equiv 0 \qquad (\text{Mod } n)$$

if, and only if,

$$x = \frac{a \cdot n}{\text{GCD}(n, k)}$$

for some integer a.

PROOF First of all, notice that if

$$x = \frac{a \cdot n}{\text{GCD}(n, k)},$$

then

$$x \cdot k = \frac{a \cdot n \cdot k}{\text{GCD}(n, k)} = a \cdot n \cdot \frac{k}{\text{GCD}(n, k)}.$$

and since $\text{GCD}(n, k)$ is a divisor of k, we see that $x \cdot k$ is a multiple of n. Thus,

$$x \cdot k \equiv 0 \ (\text{Mod } n).$$

Now suppose that $x \cdot k$ is a multiple of n. We want to show that

$$a = \frac{x \cdot \text{GCD}(n, k)}{n}$$

is in fact an integer. By the greatest common divisor theorem (1.2), there exist integers u and v such that $\text{GCD}(n, k) = u \cdot n + v \cdot k$. Then

$$a = \frac{x \cdot (u \cdot n + v \cdot k)}{n} = x \cdot u + \frac{x \cdot k \cdot v}{n}.$$

Since $x \cdot k$ is a multiple of n, we see that a is an integer. Thus,

$$x = \frac{a \cdot n}{\text{GCD}(n, k)}$$

for some integer a. □

We can now find the number of elements in a cyclic group that satisfies the equation $x^k = e$.

COROLLARY 2.1
There are precisely GCD(n, k) elements of Z_n such that $x^k = e$.

PROOF Let z be a generator of Z_n, and let $x = z^y$ be an element of Z_n. Then $x^k = (z^y)^k = z^{y \cdot k}$, which is equal to the identity if and only if

$$y \cdot k \equiv 0 \text{ (Mod } n).$$

By proposition 2.6, this is true if and only if

$$y = \frac{a \cdot n}{\text{GCD}(n, k)}$$

for some integer a. Hence, the number of possible values of y between 0 and $n - 1$ for which $z^{y \cdot k} = e$ is

$$\frac{n}{n/\text{GCD}(n, k)} = \text{GCD}(n, k).$$

Each such value of y between 0 and $n-1$ produces a different solution $x = z^y$, so there are exactly $\text{GCD}(n, k)$ solutions. □

We are now ready to consider a more complicated group. One of the puzzles that is related to the Rubik's Cube® is called the Pyraminx™. The Pyraminx™ consists of a triangular pyramid, with each of the four triangular sides partitioned into nine smaller triangles. The four "tips" can rotate, but this does not affect the puzzle. The command

ShowPuzzle

shows a simplified puzzle with the four tips chopped off, as in figure 2.3. In fact, removing the four tips gives us the advantage of being able to see the colors on the back side of the puzzle through the hole created. Now the four corners of this puzzle can rotate clockwise, using the commands

RotatePuzzle[**f**]
RotatePuzzle[**b**]
RotatePuzzle[**l**]
RotatePuzzle[**r**]

We can always put the puzzle back into its original form with the command

ResetPuzzle

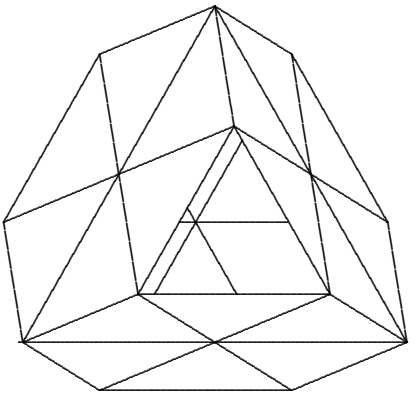


FIGURE 2.3: The Pyraminx$^{\text{TM}}$ puzzle without tips

The set of all actions on the puzzle forms a group, called the Pyraminx$^{\text{TM}}$ group. This group is generated by the elements $\{t, b, r, l\}$, and has over 900,000 elements! We can animate a sequence of moves as we did for the octahedron:

RotatePuzzle[b.f]

We can find the order of this element by repeatedly executing this command until the puzzle is back in order. In this particular case, the order of the element $b \cdot f$ is 15, meaning that we have to execute this procedure 15 times before we are back where we started.

Throughout this course, we will develop tools to work with groups that will help us to solve this puzzle, and others like it. The solution to the Pyraminx$^{\text{TM}}$, for example, is covered in section 8.4.

Problems for Chapter 2

Interactive Problems

2.1 Use *Mathematica*'s circle graph to find all of the generators of the group Z_{21}.

2.2 Use *Mathematica* or GAP to see if there an element of Z_{25}^* that generates Z_{25}^*. If so, how many such elements are there?

2.3 Use *Mathematica* or GAP to define a group that has two elements, a and b, such that $a^5 = b^4 = e$, and $b \cdot a = a^2 \cdot b$. (In GAP, use $(b \cdot a)/(a^2 \cdot b) = e$.) How many elements does this group have?

2.4 Use problem 2.3 to find the subgroup generated by the set $\{a, b^2\}$. How many elements does this subgroup have?

2.5 Use *Mathematica* to find the order of the elements $b \cdot f$, $b \cdot f \cdot r \cdot f \cdot f$, and $f \cdot b \cdot r$ in the Pyraminx$^{\text{TM}}$ group.

2.6 Can you use *Mathematica* to find an element of the Pyraminx$^{\text{TM}}$ group that has order 30?

Hint: Exactly five of the six edges must be moved out of place. The sixth edge must flip as well.

2.7 Find all of the generators of the group Z_{24}. Then have *Mathematica* or GAP construct a multiplication table for the group Z_{24}^*.

2.8 Since the elements b and c could generate the octahedral group, define this group in GAP using only b and c.

Hint: Besides $b^3 = e$ and $c^4 = e$, GAP will need one more equation. What is the order of $b^2 \cdot c$?

2.9 Define a group in GAP that is generated by two elements a and b, with $a^3 = b^5 = (a \cdot b)^2 = e$. How big is the group?

Non-Interactive Problems

For problems **2.10** through **2.12**: Find all of the generators of the following groups. How many generators are there?

2.10 Z_9^* **2.11** Z_{14}^* **2.12** Z_{18}^*

For problems **2.13** through **2.16**: Use the totient function theorem (2.1) to find the size of the following groups:

2.13 Z_{100}^* **2.14** Z_{1200}^* **2.15** Z_{1260}^* **2.16** Z_{3675}^*

2.17 Using the totient function theorem (2.1), prove that there is no value of n for which $\phi(n) = 14$.

2.18 Show that if $a^2 = b^2 = e$, then saying that $b \cdot a = a \cdot b$ is equivalent to saying that $a \cdot b \cdot a \cdot b = e$.

2.19 In defining S_3, we used three facts about the group: $a^2 = e$, $b^3 = e$, and $b \cdot a = a \cdot b^2$. Using just these facts without *Mathematica* or GAP, prove that $b^2 \cdot a = a \cdot b$.

2.20 The group defined in problem 2.3 has elements a and b such that $a^5 = e$, $b^4 = e$, and $b \cdot a = a^2 \cdot b$. Using just these facts without *Mathematica* or GAP, prove that $b^3 \cdot a = a^3 \cdot b^3$.

2.21 Write down the multiplication table for the group of symmetries of a regular tetrahedron.

Hint: Consider the octahedron with the red, yellow, orange, and cyan faces extended as to cover the other four faces. This gives us a tetrahedron, so the symmetries of a tetrahedron must be a subgroup of the octahedral group. Number the elements $1, 2, 3, \ldots, 9, T, E, W$, with 1 as the identity element. Then fill in the rest of the table. Once several elements are put in, use the Latin square property to speed up the process.

2.22 Suppose we considered rearranging four books on a shelf instead of three. How many ways could we rearrange the books?

For problems **2.23** through **2.25**, find all of the subgroups of the following groups:

2.23 Z_{12} **2.24** Z_{20} **2.25** Z_{15}^* (see table 1.4)

2.26 Use geometry to figure out how many elements of the octahedral group are of order 4. (Rotations by 90 degrees.) How many elements are of order 3? Of order 2? Check these figures by adding up these numbers, and adding one for the identity element, and show that this gives 24.

2.27 Prove that no element of the Pyraminx$^{\text{TM}}$ group can have order greater than 30.

Hint: Consider corners and edges separately. See the hint for problem 2.6.

2.28 Use corollary 2.1 to find the number of solutions to the equation $x^9 = e$ in the group Z_{18}. How many solutions are there to the equation $x^3 = e$ in this group? How many elements of order 9 are in this group?

Hint: For an element to be of order 9, it must solve $x^9 = e$, and not solve $x^n = e$ for any lower value of n.

2.29 Using only corollary 2.1, determine the number of elements of Z_{42} that are of order 6. (See the hint for problem 2.28.)

2.30 Prove that any subgroup of a finite cyclic group is cyclic.

2.31 Prove that if k is a divisor of n, then there are exactly $\phi(k)$ elements of the group Z_n that are of order k.

Hint: First do the case when $n = k$. Then use corollary 2.1 to show that the number of elements of order k for the groups Z_n and Z_k is the same.

2.32 Use problem 2.31 to show that

$$n = \sum_{k|n} \phi(k)$$

where the sum has one term for each positive divisor k of n.

2.33 If a cyclic group has an element of infinite order, how many elements of finite order does it have? Prove your answer.

2.34 Let p be a prime number. If a group G has more than $p-1$ elements of order p, prove that G cannot be a cyclic group.

2.35 Let G be an abelian group. Show that the set of elements of G that has finite order forms a subgroup of G. This subgroup is called the *torsion subgroup* of G.

2.36 Let G be an arbitrary group, with a and b two elements of G. Show that $a \cdot b$ and $b \cdot a$ have the same order.

Hint: First show by induction that $(a \cdot b)^n = a \cdot (b \cdot a)^{(n-1)} \cdot b$.

2.37 Suppose that G is a group with exactly one element of order 2, say x. Prove that $x \cdot y = y \cdot x$ for all y in G.

2.38 Let p be an odd prime number, and let $G = Z_p^*$. Show that the set

$$H = \{x^2 \quad | \quad x \in Z_p^*\}$$

forms a subgroup of G of order $(p-1)/2$. This subgroup H is called the group of *quadratic residues modulo p.*

Hint: Once you have shown that H is a subgroup, show that

$$x^2 \equiv 1 \pmod{p}$$

has exactly two solutions. Finally show that every element of H is derived from exactly two elements of Z_p^*.

This page intentionally left blank

Chapter 3

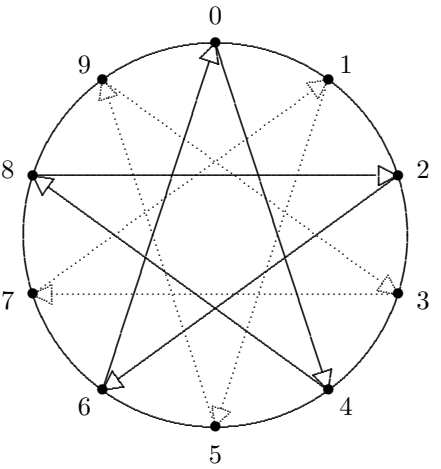
Patterns within the Cosets of Groups

3.1 Left and Right Cosets

We introduced subgroups in the last chapter, but left many questions unanswered. For example, is there any relationship between the size of the group and the size of one of its subgroups?

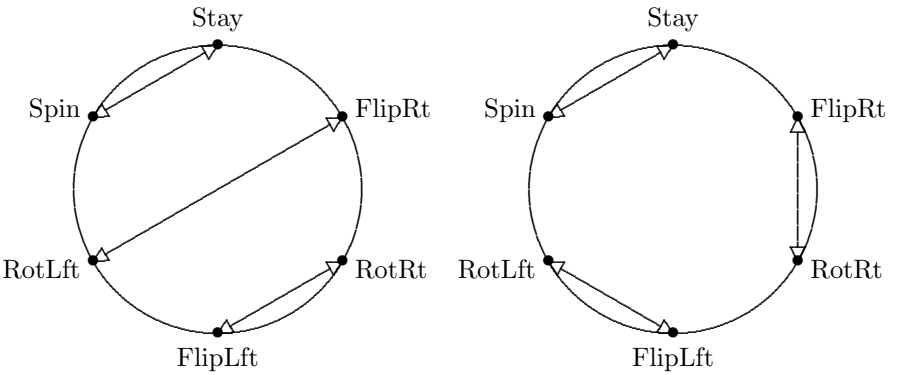
In this chapter we will introduce the tool of *cosets* to determine many of the properties of subgroups, including what possible sizes the subgroups could be. To understand cosets, let us begin by looking at some cases where an element does *not* generate the group, in hopes of finding some patterns in the circle graphs. For example, consider the element 4 from the group Z_{10}. This element does not generate the entire group, as evident from the two types of arrows in the circle graph.

DefSumMod[10]
CircleGraph[{0, 1, 2, 3, 4, 5, 6, 7, 8, 9}, Add[4]]



The solid arrows connect the points $\{0, 2, 4, 6, 8\}$, while the dotted arrows connect the points $\{1, 3, 5, 7, 9\}$. Thus, the group is partitioned into two sets, and no arrow connects these two.

One of the two sets is actually a subgroup of Z_{10}, the subgroup generated by the element 4. The other set is obtained by adding 1 to each element of the subgroup. Similar patterns arise when we use different elements of Z_{10} instead of 4.



CircleGraph[G, LeftMult[Spin]] **CircleGraph[G, RtMult[Spin]]**

FIGURE 3.1: Circle graphs showing the cosets of [Stay, Spin]

We can try a similar partitioning on non-abelian groups, such as Terry's group. If we consider forming a circle graph that sends each element to that element multiplied by **Spin**, we immediately see that we have a choice as to whether we have x map to $x \cdot \mathbf{Spin}$ or to $\mathbf{Spin} \cdot x$. The circle graph for the first option is shown in the left half of figure 3.1. This leads to a partition of the group into the sets {**Stay**, **Spin**}, {**RotRt**, **FlipLft**}, and {**RotLft**, **FlipRt**}. The latter option, shown on the right side of figure 3.1, is to multiply on the right instead of the left, giving the partition {**Stay**, **Spin**}, {**RotRt**, **FlipRt**}, and {**RotLft**, **FlipLft**}. In both cases, one of the sets in the partition is the subgroup $\mathbf{H} = \{\mathbf{Stay, Spin}\}$, but the other sets are different.

DEFINITION 3.1 Let G be a group, and let H be a subgroup of G. If x is an element of G, we define the set

$$xH = \{x \cdot y \quad | \quad y \in H\}.$$

The set xH is called a *left coset of* H. Likewise,

$$Hx = \{y \cdot x \quad | \quad y \in H\}$$

is a *right coset of* H.

Mathematica® mimics this notation. Thus,

H . RotRt

forms a right coset by multiplying every element in H by RotRt. Likewise

RotRt . H

forms a left coset. In GAP, though, we use a function `Mult` which multiplies two sets of elements from a group. The first argument gives the entire group, and the next two arguments can either be an element or a set of elements from this group.

```
gap> InitTerry();
[ Stay, FlipRt, RotRt, FlipLft, RotLft, Spin ]
gap> H := [Stay, Spin];;
gap> Mult(Terry, H, RotRt);
[ FlipRt, RotRt ]
gap> Mult(Terry, RotRt, H);
[ RotRt, FlipLft ]
```

We will denote the set of all *left* cosets of the subgroup H of G by G/H, and will denote the set of all *right* cosets of this subgroup by $H\backslash G$. Notice that the notation for right cosets uses a backward slash. In both cases, the subgroup can be considered to be on the "bottom," but since a right coset Hx has the subgroup on the left, we use $H\backslash G$, which also has H on the left, to list all such right cosets.

Mathematica and GAP find all left and right cosets of G with H with the commands

LftCoset[G, H]
and
RtCoset[G, H]

```
gap> LftCoset(Terry,H);
[ [ Stay, Spin ], [ FlipRt, RotLft ], [ RotRt, FlipLft ] ]
gap> RtCoset(Terry,H);
[ [ Stay, Spin ], [ FlipRt, RotRt ], [ FlipLft, RotLft ] ]
```

Each coset is displayed as a list of elements, so we end up with a "list of lists," giving all of the cosets.

We immediately see some patterns in the cosets. First of all, all of the cosets are the same size. Also, every element of the group appears once, and only once, in each of the two coset lists. We will prove that these patterns are true in general with two lemmas.

LEMMA 3.1

Let G be a group and H be a finite subgroup of G. Then all left and right cosets of G with respect to H contain $|H|$ elements.

PROOF It is clear from the definitions that Hu and uH each contains at most $|H|$ elements. In order to prove that the number is exactly $|H|$ we need to show that two distinct elements of H produce two different elements in the cosets. Suppose that this were not the case in a right coset. We would have two different elements x and y in H for which

$$x \cdot u = y \cdot u,$$

but multiplying on the right by u^{-1} gives $x = y$, a contradiction. Similar reasoning works for left cosets. If

$$u \cdot x = u \cdot y,$$

multiplying on the left by u^{-1} shows that $x = y$. □

Next we must show that every element of G is in exactly one left coset and one right coset. This can be worded as follows:

LEMMA 3.2
If two left or two right cosets have an element in common, they are in fact the same coset. That is,

$$Hx \cap Hy \neq \{\ \} \quad \textit{implies that} \quad Hx = Hy,$$

and

$$xH \cap yH \neq \{\ \} \quad \textit{implies that} \quad xH = yH.$$

PROOF We begin with right cosets. Suppose there is an element $g \in Hx \cap Hy$. Then there are elements h and k in H such that

$$g = h \cdot x = k \cdot y.$$

Therefore,

$$x = h^{-1} \cdot k \cdot y,$$

and so

$$(*) \qquad Hx = Hh^{-1} \cdot k \cdot y.$$

Since H is a subgroup, $h^{-1} \cdot k \in H$, so that $H \cdot h^{-1} \cdot k \subseteq H$. Moreover, if u is in H, then

$$u = (u \cdot k^{-1} \cdot h)(h^{-1} \cdot k) \in Hh^{-1} \cdot k.$$

Therefore

$$H \subseteq Hh^{-1} \cdot k,$$

and we have shown that $H = Hh^{-1} \cdot k$. Combining this with $(*)$ gives us $Hx = Hy$.

We can do left cosets in the same way. If there is an element $g \in xH \cap yH$, then there are elements h and k in H such that

$$g = x \cdot h = y \cdot k.$$

Therefore,

$$x = y \cdot k \cdot h^{-1},$$

and so

$$xH = y \cdot k \cdot h^{-1}H = yH.$$ □

With these two lemmas, we can show that the size of any subgroup is related to the size of the original group.

THEOREM 3.1: Lagrange's Theorem
Let G be a finite group, and H a subgroup of G. Then the order of H divides the order of G. That is, $|G| = k \cdot |H|$ for some positive integer k.

PROOF We can use either left cosets or right cosets to prove this, so let us use right cosets. Every element of x in G is contained in at least one right coset. For example, x is contained in Hx. Let k be the number of distinct right cosets. Then, if the right cosets are

$$Hx_1, Hx_2, Hx_3, \ldots, Hx_k,$$

we can write

$$G = Hx_1 \cup Hx_2 \cup Hx_3 \cup \cdots \cup Hx_k.$$

The $\cup$'s represent the union of the cosets. But by lemma 3.2, there are no elements in common among these sets, and so this union defines a partition of G. By lemma 3.1, each cosets contains $|H|$ elements. So $|G| = k \cdot |H|$. □

Lagrange's theorem, which seems apparent when looking at the cosets of a subgroup, turns out to have some far-reaching consequences. Let us look at some of the results that can be obtained using Lagrange's theorem.

COROLLARY 3.1
Let G be a finite group, and let x be an element of G. Then the order of x divides $|G|$.

PROOF The order of x equals the order of the subgroup $[x]$ of G. Therefore, by Lagrange's theorem (3.1), the assertion follows. □

COROLLARY 3.2
Let G be a finite group of order n and let x be an element of G. Then

$$x^n = e.$$

PROOF Let m denote the order of x. By corollary 3.1, $n = mk$ for some integer k. Then we have $x^n = x^{mk} = (x^m)^k = e^k = e$. □

COROLLARY 3.3
A group of prime order is cyclic.

PROOF Suppose G is of order p, which is prime. Then the only positive divisors of p are 1 and p, so by Lagrange's theorem (3.1) any subgroup must be of order 1 or p. If x is any element of G besides the identity, then $[x]$ contains x as well as the identity. Thus, $G = [x]$ so G is cyclic. □

COROLLARY 3.4
Let n be a positive integer, and x a number coprime to n. Then

$$x^{\phi(n)} \equiv 1 \text{ (Mod } n),$$

where $\phi(n)$ is Euler's totient function.

PROOF We simply apply corollary 3.2 to the group Z_n^*. This group has $\phi(n)$ elements, and if x is coprime to n then x is a generator of Z_n, so x is in Z_n^*. □

In particular, when $n = p$ is prime, we have

$$x^{p-1} \equiv 1 \text{ (Mod } p).$$

This result is known as Fermat's little theorem.

DEFINITION 3.2 If H is a subgroup of G, we define the *index* of H in G, denoted $[G\colon H]$, to be the number of right cosets in $H\backslash G$. Of course this is the same as the number of left cosets in G/H.

Notice that when G is a finite group we have by the argument in Lagrange's theorem (3.1) that $|G| = |H| \cdot [G\colon H]$.

3.2 How to Write a Secret Message

It was mentioned in the last section that Lagrange's theorem (3.1) has some far-reaching implications. One of these implications is the ability to write a message that no one can read except for the person to whom the message is sent, *even if the whole world knows the code!*

To introduce this code, we begin by considering the group Z_{33}^*, whose order is $\phi(33) = 20$. The elements of Z_{33}^* are

$$\{1, 2, 4, 5, 7, 8, 10, 13, 14, 16, 17, 19, 20, 23, 25, 26, 28, 29, 31, 32\}.$$

Consider the mapping that sends every element to its square. In essence we are defining a function $f(x) = x^2$ on this group. We can make a circle graph in *Mathematica* that maps each element to its square by the command

DefMultMod[33]
CircleGraph[{1,2,4,5,7,8,10,13,14,16,17,19,20,23,25,26,28,29,31,32}, Pow[2]]

which produces figure 3.2.

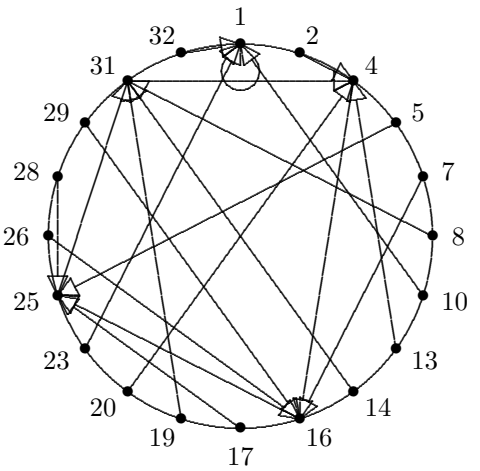


FIGURE 3.2: Circle graph for squaring in Z_{33}^*

This graph is rather perplexing. The squares of 2, 13, 20, and 31 are all 4. The elements having "square roots" have four of them, while the majority of the elements do not have square roots.

If we try cubing each element instead, using the command

CircleGraph[{1,2,4,5,7,8,10,13,14,16,17,19,20,23,25,26,28,29,31,32}, Pow[3]]

we get figure 3.3. This graph has a very different behavior: no two elements have the same cube. Also, every element has a "cube root." The terminology used for standard functions over the real numbers can be used for functions defined on groups.

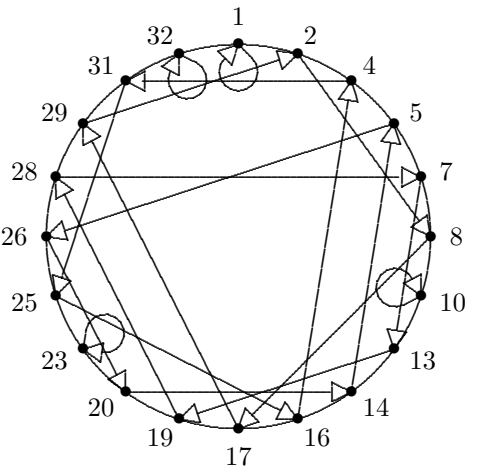


FIGURE 3.3: Circle graph for cubing in Z_{33}^*

DEFINITION 3.3 We say that a function $f(x)$ is *one-to-one* if the only way for $f(x) = f(y)$ is for $x = y$. We say that a function $f(x)$ is *onto* if for every y, there is an x such that $f(x) = y$.

In terms of the circle graphs, a one-to-one function cannot have two arrows pointing to the same point. Likewise, an onto function will have at least one arrowhead at each point. We see from figure 3.3 that the cube function is both one-to-one and onto. Thus, every element has a unique cube root.

In fact, the cube root of any element in this group can be found by taking the seventh power of the element! This is because $\phi(33) = 20$, so using corollary 3.4,

$$(x^3)^7 = x^{21} = x^{20} \cdot x = e \cdot x = x.$$

The key difference between the squaring function and the cubing function stems from the fact that 3 is coprime to $\phi(33) = 20$, whereas 2 is not.

PROPOSITION 3.1

Suppose G is a finite group of order m, and that r is some integer which is coprime to m. Then the function $f(x) = x^r$ is one-to-one and onto. In other words, we can always find the unique r-th root of any element in G.

PROOF Since G is of order m, we have by corollary 3.2 that $x^m = e$ for all x in G. If r and m are coprime, then r is a generator in the additive group Z_m. But this means that r is an element of the group Z_m^*, and so there is an inverse element $s = r^{-1}$. Thus, $s \cdot r \equiv 1$ in Z_m^*. Another way we could say

this is

$$sr = km + 1$$

for some integer k.

Now we are ready to take the r-th root of a number. If y is an element of G, then the r-th root of y in G is merely y^s. To see this, note that

$$(y^s)^r = y^{sr} = y^{(km+1)} = (y^m)^k \cdot y = e^k \cdot y = y.$$

So y^s is one r-th root of a. But y^s must be a different element for every y in G, since the r-th power of y^s is different. Since the r-th root of every element of G is accounted for, by the pigeonhole principle there cannot be *two* r-th roots to any element. Thus, y^s gives the unique r-th root of y in G. □

Let us now consider the cubes of all numbers from 0 to 32. This will no longer be a group, since we have included non-invertible elements. But with the circle graph shown in figure 3.4, we find that the mapping $x \to x^3$ is still one-to-one and onto. Thus, we can still find the cube root of a number modulo 33 by taking the seventh power modulo 33. The reason is given in the next proposition.

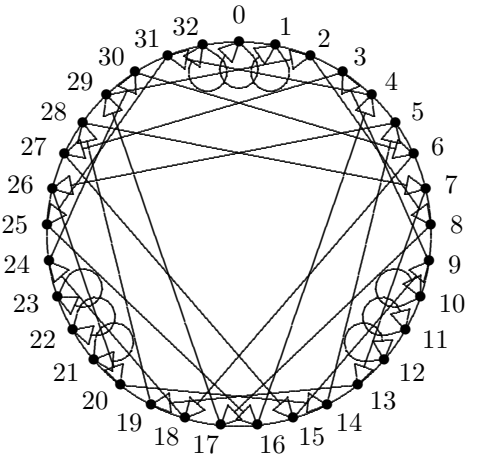


FIGURE 3.4: Circle graph for cubing modulo 33

PROPOSITION 3.2

Suppose n is a product of two distinct primes and

$$r \cdot s \equiv 1 \text{ (Mod } \phi(n)).$$

Then for all values of x less then n,

$$(x^r)^s \equiv x \text{ (Mod } n).$$

PROOF If x is coprime to n, then proposition is true by proposition 3.1. Suppose x is not coprime to $n = p \cdot q$, where p and q are the two distinct primes. By the totient function theorem (2.1), $\phi(n) = (p-1) \cdot (q-1)$. The number x would be a multiple of either p or q, say p. Then

$$x^{r \cdot s} = (p \cdot a)^{r \cdot s} = p^{r \cdot s} \cdot a^{r \cdot s}$$

will be a multiple of p. Also, x is *not* a multiple of q since x is less than n. Since $r \cdot s \equiv 1 \pmod{(p-1)(q-1)}$, $r \cdot s \equiv 1 \pmod{(q-1)}$. Thus, by proposition 3.1 again, we have

$$x^{rs} \equiv x \pmod{q}.$$

Since we also have $x^{rs} \equiv x \pmod{p}$, by the Chinese Remainder Theorem (1.3), we have, since p and q are coprime,

$$x^{rs} \equiv x \pmod{pq = n}.$$

□

The function $x \to x^3$ is not only one-to-one and onto, but also mixes up the numbers 0 through 32 fairly well. This suggests an encryption scheme. We can first convert a message to a sequence of numbers using table 3.1. For example,

CAN YOU READ THIS

becomes

3, 1, 14, 0, 25, 15, 21, 0, 18, 5, 1, 4, 0, 20, 8, 9, 19.

The encryption scheme is to replace each number with its cube, modulo 33.

TABLE 3.1: Standard code sending letters to numbers

A	=	1	J	=	10	S	=	19
B	=	2	K	=	11	T	=	20
C	=	3	L	=	12	U	=	21
D	=	4	M	=	13	V	=	22
E	=	5	N	=	14	W	=	23
F	=	6	O	=	15	X	=	24
G	=	7	P	=	16	Y	=	25
H	=	8	Q	=	17	Z	=	26
I	=	9	R	=	18	Space	=	0.

This gives us

27, 1, 5, 0, 16, 9, 21, 0, 24, 26, 1, 31, 0, 14, 17, 3, 28.

To decipher this, one would take the seventh power of each number in the sequence modulo 33, and convert back to letters in the alphabet.

The main drawback with this code is that, for longer messages, the letter E which encodes to 26 would appear most frequently in the encoded string. Someone who didn't know the code might deduce that 26 stands for E without knowing anything about algebra. But also anyone who knew how to encrypt the message could use proposition 3.1 to decipher the message, for they could deduce that 7 is the inverse of 3 modulo 20. What we need is a code in which everyone would know how to encrypt a message, but only the person who originated the code could decipher.

We can solve both of these problems just by picking n to be the product of two huge prime numbers p and q, say 80 digits each. Then $\phi(n) = (p-1) \cdot (q-1)$. We then pick r to be a number of at least four digits that is coprime to $\phi(n)$. The encryption scheme is then

$$x \to y = x^r \text{ (Mod } n).$$

We decode this by finding $s = r^{-1}$ in the group $Z^*_{\phi(n)}$. By proposition 3.2, the operation

$$y \to x = y^s \text{ (Mod } n)$$

"undoes" the encryption, since

$$(x^r)^s = x \text{ (Mod } n).$$

One big advantage of using huge numbers for the code is that we can encrypt an entire *line* at a time. For example,

CAN YOU READ THIS

can be encrypted by the single number

030114002515210018050104002008091 9

by having every two digits represent one letter (still using table 3.1). This prevents cracking the code using the frequencies of the letters. But the unusual advantage of this code is that only the originator of the code can decipher a message, even if the encryption scheme and the values of n and r were made public.

In order to decode a message, one must know the value of s, which is given by the inverse of r (Mod $\phi(n)$). This is easy to do with *Mathematica* or GAP once $\phi(n)$ is known, but how difficult it is to find $\phi(n)$! One needs to know the prime factorization of n, which would be about 160 digits long. Even GAP or *Mathematica* could not factor this in a reasonable amount of time. In fact, adding two digits to p and q makes the factorization 10 times harder. So by making the prime numbers larger, we can be assured that the factorization cannot be done within one's lifetime. [6, p. 21] Thus, without knowing the

original primes p and q that were multiplied together, it is virtually impossible to determine s.

This encryption scheme is called the Rivest-Shamir-Adleman encryption.[6, p. 374] Both *Mathematica* and GAP have built in routines that allow us to experiment with RSA encryption. The *Mathematica* function

p = NextPrime[12345678901234567890123456789012345678901234567890123456789012345678901234567890]

finds the next prime number larger than that 80 digit number. In GAP, the corresponding function is `NextPrimeInt`. Since we want n to be the product of two large primes, we will find another large prime q, and multiply these primes together.

```
gap> p := NextPrimeInt(12345678901234567890123456789012345678901234567890);;
#I  IsPrimeInt: probably prime, but not proven:
12345678901234567890123456789012345678901234567890123456789012345678901234567997
gap> q := NextPrimeInt(98765432109876543210987654321098765432109876543210);;
#I  IsPrimeInt: probably prime, but not proven:
98765432109876543210987654321098765432109876543210987654321098765432109876543391
gap> n := p*q;;
```

In both GAP and *Mathematica*, we can use a backslash to break the input into two lines, and it will be read as a single line. GAP issues a warning that these numbers are only probably prime, but the odds of a non-prime number passing the prime test are astronomically small, so we can safely assume that these are indeed prime. This is true in *Mathematica* as well, but no warning is issued. In *Mathematica*, we finish this up with the commands

q = NextPrime[98765432109876543210987654321098765432109876543210987654321098765432109876543210]
n = p q

The number n can be made public, along with any four digit number r that is coprime to both $p-1$ and $q-1$. For simplicity, we will use a four digit prime number.

r = NextPrime[1234]

```
gap> r := NextPrimeInt(1234);
1237
```

We can verify that this is coprime to $(p-1)(q-1)$ by computing

GCD[(p–1)(q–1), r]

or

```
gap> GcdInt((p-1)*(q-1), r);
1
```

which returns 1.

To encrypt a message, the command

x = MessageToNumber["HERE IS A MESSAGE"]

converts any sentence into a number. Note that the message is put in quotation marks. This number can now be encrypted by the command

y = PowerMod[x, r, n]

In GAP, we use `PowerModInt` instead of **PowerMod**.

```
gap> x := MessageToNumber("HERE IS A MESSAGE");
8051805000919000100130519190107O5
```

```
gap> y := PowerModInt(x, r, n);
14724730500997597506102032344396082021733211823548530129332813\
79106660097841745903879602610137146145206880730757815860390004\
76825576155377145604282754058969344
```

Deciphering a message is very similar, only we will use the secret number s instead of r. Suppose a friend, knowing the values of n and r, gives the message

**y = 69557405147024406870611426657425604382775606544074705\
32387700788446830783525388331288538827113160595765080505\
96669314319991863521509357081622413906361655183O794**

```
gap> y := 69557405147024406870611426657425604382775606544074705\
>     32387700788446830783525388331288538827113160595765080505\
>     96669314319991863521509357081622413906361655183O794;;
```

To decode the message, we first need to know the value of s, which is the inverse of r modulo $(p-1)(q-1)$. Thus, the command to find s is given by

s = PowerMod[r, –1, (p–1)(q–1)]

```
gap> s := PowerModInt(r, -1, (p-1)*(q-1));;
```

Next, compute y^s (Mod n) by the command

x = PowerMod[y, s, n]

Finally, the command

NumberToMessage[x]

puts the message into readable form. In GAP, these final steps are as follows:

```
gap> x := PowerModInt(y, s, n);
135555700063550051700037403330006693639300525558596454007058555\
006958555493
gap> NumberToMessage(x);
"Meet me at 7:30 p.m. behind the shed."
```

You may notice that the encryption in table 3.1 has been expanded to allow lower case letters and punctuation. There are many other applications to this code besides sending secret messages. For example, suppose to get an account at the Electronic Bank, you pick two large random prime numbers, p and q. The bank then gives you the account number $n = p \cdot q$, and a number r, and makes these public. The bank also gives you the secret number

$$s = r^{-1} \text{ (Mod } (p-1)(q-1)).$$

You use the number s to *decode* messages such as

MessageToNumber[
"Check 1034: Pay to the order of John Brown $43.50"]
x = PowerMod[%, **s, n**]

```
gap> MessageToNumber(
> "Check 1034:  Pay to the order of John Brown $43.50");
358555361003130333440000016517500706500705855006568545568006555\
600106558640002686573640083343393353O
gap> x := PowerModInt(last, s, n);
759886203333804191757867804397582340158888583830837689727777591\
850158788227670494169489490389712206354728907657364155336042701\
750568998247000003691863304794999l8
```

This number, along with your account number and the number r, is sent to John Brown. His bank can verify that this number is in fact a check as follows:

y = PowerMod[**x, r, n**]
NumberToMessage[**y**]

```
gap> y := PowerModInt(x, r, n);;
gap> NumberToMessage(y);
"Check 1034:  Pay to the order of John Brown $43.50"
```

This proves that the only person knowing s sent this message. Hence, the encryption acts as a *signature* to the check. Using this method, one can send an "electronic check" (even through e-mail) that is virtually impossible to forge.

3.3 Normal Subgroups

We can define a product of any *subset* of a group G by an element of G in the same way that we defined a product of a subgroup and an element. That

is, if X is any subset of G, we can define

$$\begin{aligned} Xu &= \{x \cdot u \quad | \quad x \in X\}, \qquad \text{and} \\ uX &= \{u \cdot x \quad | \quad x \in X\}. \end{aligned}$$

If X and Y are two subsets of a group G, we can also define

$$X \cdot Y = \{x \cdot y \quad | \quad x \in X \text{ and } y \in Y\}$$

By defining the product of subsets in this way, we find that $\{u\} \cdot X = uX$. We also discover that

$$X \cdot (Y \cdot Z) = (X \cdot Y) \cdot Z.$$

This raises some interesting questions. If X and Y are subgroups of G, will $X \cdot Y$ be a subgroup? Suppose X and Y are cosets of G with respect to a subgroup H. Will $X \cdot Y$ be a coset of G?

We will use the octahedral group of order 24 to experiment. In *Mathematica*, this can be reloaded with the commands

InitGroup[e]; Define[a^2, e]; Define[b^3, e]; Define[c^4, e]
Define[1/a, a]; Define[1/b, b^2]; Define[1/c, c^3]
Define[b.a, a.b.b]; Define[c.a, a.b.c]; Define[c.b, a.c.c]
G = Group[{a, b, c}];

Two sample subgroups of order 4 are given by

H = Group[{c}]
$\{e,\ c,\ c \cdot c,\ c \cdot c \cdot c\}$

and

K = Group[{b.c}]
$\{e,\ b \cdot c,\ a \cdot b \cdot c \cdot c,\ a \cdot b \cdot b \cdot c \cdot c \cdot c\}$

whose product can be computed using *Mathematica*.

H . K
$\{c,\ e,\ a \cdot b,\ a \cdot c,\ b \cdot b,\ b \cdot c,\ c \cdot c,\ a \cdot b \cdot b,\ c \cdot c \cdot c,\ a \cdot b \cdot b \cdot c,\ a \cdot b \cdot c \cdot c,$
$a \cdot c \cdot c \cdot c,\ b \cdot b \cdot c \cdot c,\ b \cdot c \cdot c \cdot c,\ a \cdot b \cdot b \cdot c \cdot c,\ a \cdot b \cdot b \cdot c \cdot c \cdot c\}$

In GAP, the commands are

```
gap> f:=FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> g:=f/[a^2,b^3,c^4, b*a/(a*b*b), c*a/(a*b*c), c*b/(a*c*c)];;
gap> a:=g.1;; b:=g.2;; c:=g.3;;
gap> G := ListGroup(g);
[ <identity ...>, a, b, a*b, b^2, a*b^2, c, a*c, b*c, a*b*c,
  b^2*c, a*b^2*c, c^2, a*c^2, b*c^2, a*b*c^2, b^2*c^2,
  a*b^2*c^2, c^3, a*c^3, b*c^3, a*b*c^3, b^2*c^3, a*b^2*c^3 ]
```

```
gap> H := Group(c);
Group([ c ])
gap> K := Group(b*c);
Group([ b*c ])
gap> Mult(G,H,K);
[ <identity ...>, a*b^2, a*b^2*c, c^2, c, a*b^2*c^2, b^2,
  a*b^2*c^3, a*b, c^3, a*c, b^2*c^2, b*c, a*b*c^2, a*c^3,
  b*c^3 ]
```

Both GAP and *Mathematica* show a set containing 16 elements. This cannot be a subgroup by Lagrange's theorem (3.1), since 16 is not a factor of 24. Note that by having G be the `ListGroup` of the group in GAP, the elements are displayed the way *Mathematica* displays the elements. This causes GAP's output for many operations to match *Mathematica*'s output.

If we consider multiplying two right cosets of H, say the third and the fifth,

```
gap> R := RtCoset(G,H);
[ [ <identity ...>, c^2, c, c^3 ],
  [ a*b^2, a*b^2*c, a*b^2*c^2, a*b^2*c^3 ],
  [ b, b^2*c, a*c^2, a*b*c^3 ], [ a, a*b*c, b*c^2, b^2*c^3 ],
  [ b^2, a*c, a*b*c^2, b*c^3 ], [ a*b, b^2*c^2, b*c, a*c^3 ] ]
gap> Mult(G, R[3], R[5]);
[ <identity ...>, a*b^2, b, a*b^2*c, c^2, a, c, a*b^2*c^2,
  a*b*c, b*c^2, a*b^2*c^3, b^2*c, a*c^2, c^3, a*b*c^3,
  b^2*c^3 ]
```

we get something equally fruitless. However, a left coset multiplied by a right coset produces a glimmer of hope:

```
gap> L := LftCoset(G,H);
[ [ <identity ...>, c^2, c, c^3 ],
  [ a*b^2, a*b^2*c, a*b^2*c^2, a*b^2*c^3 ],
  [ b, b*c^2, b*c, b*c^3 ], [ a, a*c^2, a*c, a*c^3 ],
  [ b^2, b^2*c, b^2*c^2, b^2*c^3 ],
  [ a*b*c, a*b, a*b*c^3, a*b*c^2 ] ]
gap> Mult(G, L[3], R[5]);
[ <identity ...>, a*b^2*c, a*c^2, b^2*c^3 ]
```

which a `MultTable` command shows is indeed a subgroup. In fact, experimenting shows that any left coset in L times a right coset in R will give four elements, which looks like some sort of coset.

So what happens if we find a subgroup for which the right cosets and the left cosets are the same? Then the product of a left coset and a right coset would merely be the product of two cosets. An example of such a subset is

M = {e, c.c, a.b.b.c, a.b.b.c.c.c}

which we can verify in *Mathematica* by the commands

R = RtCoset[G, M]
L = LftCoset[G, M]

or in GAP as follows:

```
gap> M := Group(c^2, a*b^2*c);
Group([ c^2, a*b^2*c ])
gap> R := RtCoset(G,M);
[ [ <identity ...>, a*b^2*c, c^2, a*b^2*c^3 ],
  [ a*b^2, c, a*b^2*c^2, c^3 ], [ b, a*b*c, b*c^2, a*b*c^3 ],
  [ a, b^2*c, a*c^2, b^2*c^3 ], [ b^2, a*c, b^2*c^2, a*c^3 ],
  [ a*b, b*c, a*b*c^2, b*c^3 ] ]
gap> L := LftCoset(G,M);
[ [ <identity ...>, a*b^2*c, c^2, a*b^2*c^3 ],
  [ a*b^2, c, a*b^2*c^2, c^3 ], [ b, a*b*c, b*c^2, a*b*c^3 ],
  [ a, b^2*c, a*c^2, b^2*c^3 ], [ b^2, a*c, b^2*c^2, a*c^3 ],
  [ a*b, b*c, a*b*c^2, b*c^3 ] ]
```

Two of these cosets are

$\mathbf{H} = \{\mathbf{a, a.c.c, b.b.c, b.b.c.c.c}\}$
$\mathbf{K} = \{\mathbf{b, a.b.c, b.c.c, a.b.c.c.c}\}$

and the product $H \cdot K$ turns out to be another coset. In fact, the product of any two cosets of the subgroup M will yield a coset of M.

```
gap> Mult(G,R[4],R[3]);
[ a*b, b*c, a*b*c^2, b*c^3 ]
```

First, let us give some terminology for this special type of subgroup.

DEFINITION 3.4 A subgroup H of the group G is said to be *normal* if all left cosets are also right cosets, and conversely, all right cosets are also left cosets. That is, H is normal if $G/H = H\backslash G$.

Next, we need a way to test whether a subset is normal.

PROPOSITION 3.3
A subgroup H is a normal subgroup of G if, and only if, $uHu^{-1} = H$ for all elements u in G.

PROOF First of all, suppose H is normal, and let u be an element of G. Then uH and Hu both contain the element u. Since the left and right cosets are the same, we have

$$uH = Hu.$$

Multiplying both sides on the right by u^{-1} gives

$$uHu^{-1} = Hu \cdot u^{-1} = H.$$

Now, suppose that $uHu^{-1} = H$ for all elements u in G. Then

$$Hu = (uHu^{-1}) \cdot u = uHe = uH.$$

Thus, every left coset is also a right coset, and vice versa. □

This gives us a way to determine if a subgroup is normal, but we can improve on this test.

PROPOSITION 3.4
Let H be a subgroup of G. Then H is normal if, and only if,

$$uHu^{-1} \subseteq H$$

for all elements $u \in G$.

PROOF The "only if" part of this statement is obvious from proposition 3.3. So let us suppose that for all u in G,

$$uHu^{-1} \subseteq H.$$

However, since $(u^{-1})^{-1} = u$, we have

$$u^{-1}Hu = u^{-1}H(u^{-1})^{-1} \subseteq H.$$

Multiplying every element in the set by u on the left gives us $Hu \subseteq uH$, and multiplying on the right by u^{-1} gives us $H \subseteq uHu^{-1}$. Since we also have that $uHu^{-1} \subseteq H$, we can conclude that $uHu^{-1} = H$. Then from proposition 3.3, H is normal. □

Thus, to test whether H is a normal subgroup, we simply have to show that $g \cdot h \cdot g^{-1}$ is in H whenever $h \in H$ and $u \in G$. There are many other examples of normal subgroups. For example, if G is any group, then the subgroups $\{e\}$ and G are automatically normal. These normal subgroups are said to be *trivial.* If G is commutative, then any subgroup will be a normal subgroup. Here is another way to tell a subgroup is normal.

PROPOSITION 3.5
If H is a subgroup of G with index 2, then H is a normal subgroup.

PROOF Since H is a subgroup of G with index 2, there are two left cosets and two right cosets. One of the left cosets is eH, which is the set of elements in H. The other left coset must then be the set of elements not in H. But the same thing is true for the right cosets, so the left and right cosets are the same. Thus, H is normal. □

When we have a normal subgroup, the set of cosets will possess more properties than for standard subgroups. We will explore these in the next section.

3.4 Quotient Groups

In the last section we observed a case where H was a normal subgroup of G, and the product of two cosets yielded another coset. Let us begin by proving that this will always happen for normal subgroups.

LEMMA 3.3
If N is a normal subgroup of G, then the product of two cosets of N is again a coset of N. In fact,

$$aN \cdot bN = (a \cdot b)N.$$

PROOF We simply observe that

$$aN \cdot bN = a \cdot (Nb) \cdot N = a \cdot (bN) \cdot N = (a \cdot b) \cdot (N \cdot N) = (a \cdot b)N.$$

Note that $Nb = bN$ because N is a normal subgroup. □

This result is very suggestive. If we can multiply two cosets to produce another coset, will the set of all cosets form a group?

THEOREM 3.2: The Quotient Group Theorem
Let N be a normal subgroup of G. Then the set of all cosets is a group, which is denoted by G/N, called the quotient group *of G with respect to N.*

PROOF We simply have to check that G/N satisfies the four requirements in definition 1.3. The closure property is given by lemma 3.3. To check associativity,

$$\begin{aligned} aN \cdot (bN \cdot cN) &= aN \cdot (b \cdot c)N = (a \cdot (b \cdot c))N \\ &= ((a \cdot b) \cdot c)N = (a \cdot b)N \cdot cN = (aN \cdot bN) \cdot cN. \end{aligned}$$

The identity element is $eN = N$, and we can check that

$$\begin{aligned} eN \cdot aN &= (e \cdot a)N = aN, \qquad \text{and} \\ aN \cdot eN &= (a \cdot e)N = aN. \end{aligned}$$

Finally, the inverse of aN is $a^{-1}N$, since

$$\begin{aligned} aN \cdot a^{-1}N &= (a \cdot a^{-1})N = eN = N, \qquad \text{and} \\ a^{-1}N \cdot aN &= (a^{-1} \cdot a)N = eN = N. \end{aligned}$$

Thus, the set of all cosets forms a group. □

One of the easiest groups to consider is the group of integers $\mathbb{Z}$ under addition. A subgroup of $\mathbb{Z}$ would consist of all multiples of k, with $k \geq 0$. ($k = 0$ and $k = 1$ produce the two trivial subgroups.) We will denote this normal subgroup of $\mathbb{Z}$ by $k\mathbb{Z}$. All elements in each coset would be equivalent modulo k. Thus, there would be k cosets of $k\mathbb{Z}$ (except when $k = 0$). Hence, $\mathbb{Z}/k\mathbb{Z}$ is essentially the same group as Z_k. The notation

$$x \equiv y \text{ (Mod } k)$$

indicates that x and y belong to the same coset of the subgroup $k\mathbb{Z}$.

We can extend this notation to any normal subgroup. We say that

$$x \equiv y \text{ (Mod } N)$$

to indicate x and y belong in the same coset of G with respect to N. It is easy to see that

$$x \equiv y \text{ (Mod } N) \quad \text{if, and only if,} \quad x \cdot y^{-1} \in N.$$

The partitioning of the cosets makes it obvious that equivalence (Mod N) satisfies the following three properties:

1. (Reflexive) Every element x is equivalent to itself.
2. (Symmetric) If x is equivalent to y, then y is equivalent to x.
3. (Transitive) If x is equivalent to y, and y in turn is equivalent to z, then x is equivalent to z.

DEFINITION 3.5 Any relationship that satisfies these three properties is called an *equivalence relationship.*

Any equivalence naturally divides a set up into smaller subsets, where members of each subset are equivalent to each other. These subsets are called *equivalence classes.*

In the last section we found a normal subgroup of the octahedral group, namely

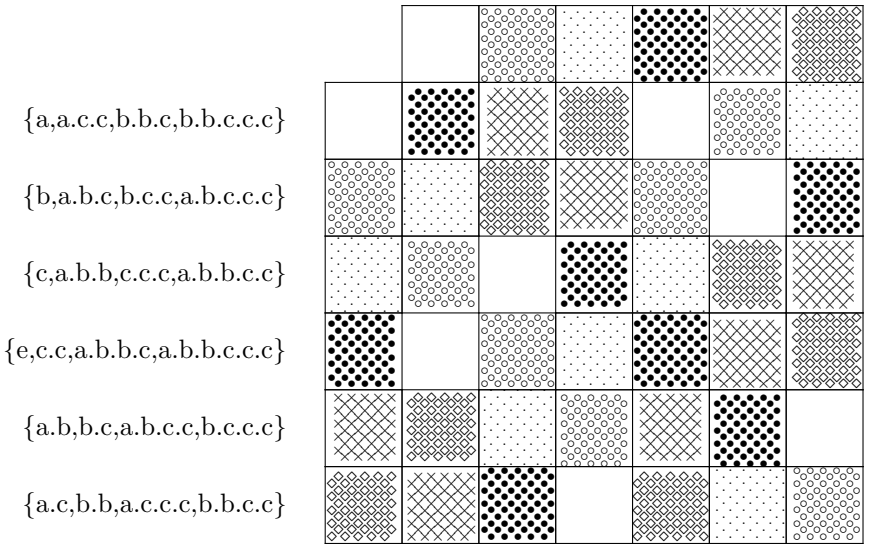
M = {e, c.c, a.b.b.c, a.b.b.c.c.c}

The cosets, or equivalence classes, with respect to this subgroup are given by the command

Q = LftCoset[G, M]

We can use *Mathematica* to give us a multiplication table of the quotient group Q.

MultTable[Q]



Since the names of the elements are so long, *Mathematica* uses a color code for the elements, which is shown here as shading. Notice that this table is very similar to the table for the group S_3, but is not quite the same color pattern, since the identity element of Q is not listed first. If we do these calculations in GAP, we do not have this problem. Note: If the group is still loaded from the last section, we can skip to the `Q := RtCoset(G,M);` command.

```
gap> f:=FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> g:=f/[a^2, b^3, c^4,b*a/(a*b*b),c*a/(a*b*c),c*b/(a*c*c)];;
gap> a:=g.1;; b:=g.2;; c:=g.3;;
gap> G := ListGroup(g);
[ <identity ...>, a, b, a*b, b^2, a*b^2, c, a*c, b*c, a*b*c,
  b^2*c, a*b^2*c, c^2, a*c^2, b*c^2, a*b*c^2, b^2*c^2,
  a*b^2*c^2, c^3, a*c^3, b*c^3, a*b*c^3, b^2*c^3, a*b^2*c^3 ]
gap> M := Group(c^2, a*b^2*c);;
gap> Q := RtCoset(G,M);
[ [ <identity ...>, a*b^2*c, c^2, a*b^2*c^3 ],
  [ a*b^2, c, a*b^2*c^2, c^3 ], [ b, a*b*c, b*c^2, a*b*c^3 ],
  [ a, b^2*c, a*c^2, b^2*c^3 ], [ b^2, a*c, b^2*c^2, a*c^3 ],
  [ a*b, b*c, a*b*c^2, b*c^3 ] ]
gap> NumberElements := true;;
gap> MultTable(Q);
*                        |1  2  3  4  5  6
-------------------------+------------------
[e,a*b^2*c,c^2,a*b^2*c^3]|1  2  3  4  5  6
[a*b^2,c,a*b^2*c^2,c^3]  |2  1  4  3  6  5
[b,a*b*c,b*c^2,a*b*c^3]  |3  6  5  2  1  4
[a,b^2*c,a*c^2,b^2*c^3]  |4  5  6  1  2  3
[b^2,a*c,b^2*c^2,a*c^3]  |5  4  1  6  3  2
[a*b,b*c,a*b*c^2,b*c^3]  |6  3  2  5  4  1
```

The command `NumberElements := true;;` allows larger tables to be displayed by substituting a number for each element. Since the order of the elements along the top is the same as the order in the leading column, it is easy to determine which element (coset) corresponds to which number. The group S_3 is already defined in *Mathematica* and GAP as a subset of the octahedral group. This subset is given by

H = {e, a, b, a.b, b.b, a.b.b}

Thus, the multiplication table of S_3 is given by

```
gap> e := Identity(g);
<identity ...>
gap> H := [e, a, b, a*b, b^2, a*b^2];;
gap> MultTable(H);

*    |1  2  3  4  5  6
-----+------------------
e    |1  2  3  4  5  6
a    |2  1  4  3  6  5
b    |3  6  5  2  1  4
a*b  |4  5  6  1  2  3
b^2  |5  4  1  6  3  2
a*b^2|6  3  2  5  4  1
```

With this particular arrangement of the elements, we see that the number patterns for Q and H match. In chapter 4, we will define two groups that have the same number or color pattern as being *isomorphic.*

Problems for Chapter 3

Interactive Problems

3.1 This exercise is required in order to do the RSA encryption problem 3.2 or 3.3. Using *Mathematica*'s **NextPrime** command, or GAP's `NextPrimeInt`, find two large prime numbers p and q, at least 80 digits each. This is done by the two *Mathematica* commands

p = NextPrime[***large number goes here*** **]**
q = NextPrime[***another large number goes here*** **]**

or the GAP commands

```
gap> p := NextPrimeInt( large number goes here );;
gap> q := NextPrimeInt( another large number goes here );;
```

We will use the value $r = 10007$. Verify that this number is coprime to $p-1$ and $q-1$ by executing the following:

GCD[(p–1)(q–1), 10007]

or

```
gap> GcdInt( (p-1)*(q-1), 10007);
```

If this yields 10007 instead of 1, go back and find new values for p and q. Once the GCD is 1, compute $n = p \cdot q$, and save this on a thumb drive. To do this, place your thumb drive in the computer (say it becomes the E: drive) and enter:

n = p q
Save["E:/nfile", n]

or in GAP,

```
gap> n := p*q;;
gap> PrintTo("E:/nfile", "n:=",n,";");
```

Note: If the thumb drive is some other drive, such as the F: drive, you will have to replace the E: with F: in the last statement, and also the statements below. Next, find the secret number s, which deciphers a message:

s = PowerMod[10007, –1, (p–1)(q–1)]

or

```
gap> s := PowerModInt(10007, -1,(p-1)*(q-1));
```

You will want to save this number for future reference. With your thumb drive still in the computer, enter

Save["E:/secret", s]

or

```
gap> PrintTo("E:/secret","s:=",s,";");
```

This number will be needed for future assignments. Don't lose it! Finally, e-mail the "nfile" file as an attachment to the professor. Alternatively, you can cut and paste the contents of "nfile" into the body of the message. Do not send the contents of the secret file.

3.2 Using the values of n and s from problem 3.1, send an "electronic check" to your favorite professor for $100.00. This check will be in the form of a huge number, x. Once this number is found, insert your thumb drive and enter

Save["E:/check", x]

or

```
gap> PrintTo("E:/check","x:=",x,";");
```

E-mail the file "check" as a file attachment, or cut and paste the contents of the file into the body of a letter.

3.3 After doing problem 3.1, you will receive a response with an attachment file "message." Save this to your thumb drive and enter

<<E:/message
<<E:/nfile
<<E:/secret

or in GAP,

```
gap> Read("E:/message");
gap> Read("E:/nfile");
gap> Read("E:/secret");
```

The first command sets y to the encrypted message, while the second command reads in your value of n. The third command loads the secret number into s that you were asked to save in problem 3.1. Using this value of s, decode the message and hand in (on paper) what it says.

3.4 B. L. User tried creating his encryption number with the two primes

p = NextPrime[715870273457197548734156715678567821637 4
15615197371557525256736492867395847560 92]
q = NextPrime[p+1]

or, in GAP,

```
gap> p := NextPrimeInt(7158702734571975487341567156785678216374\
> 15615197371557525256736492867395847560 92);;
gap> q := NextPrimeInt(p+1);;
```

When he publicized the product $n = pq$, along with the value $r = 6367$, he received a message from a friend:

y = 309272252199306433540387847641451588319943220486905800597614 0\
72507354652310684824949153128245664045438567847210761652124 20\
43590910817888839981759972041752306977

What did this message say?

3.5 Show that there is a group Q which is generated by two elements a and b, for which

$$a^4 = e, \qquad b^2 = a^2, \qquad b \cdot a = a^3 \cdot b, \qquad a^2 \neq e.$$

This can be entered into *Mathematica* with the command

InitGroup[e];
Define[a^4, e]
Define[b^2, a^2]
Define[b.a, a.a.a.b]
Q = Group[{a, b}]

or in GAP by the commands

```
gap> f := FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g := f/[a^4,(b^2)/(a^2),(b*a)/(a^3*b)];;
gap> a := g.1;; b := g.2;;
gap> Q := List(g);
```

Find all subgroups of this group, and show that all subgroups are normal, even though the group is non-abelian. (Write down the list of left cosets and right cosets for each subgroup found.)

3.6 Define $G = Z_{105}^*$ in *Mathematica*. How many elements does this group have? Consider the subgroup H generated by the element 11. A circle graph demonstrating the cosets G/H can be obtained by the command

CircleGraph[G, Mult[11]]

By looking at the circle graph, determine the cosets of G with respect to H. What is the order of the element $2 \cdot H$ in the quotient group G/H?

3.7 Use *Mathematica* or GAP, along with a bit of trial and error, to find a subgroup of order 12 of the octahedral group. Show that this subgroup is a normal subgroup. The following reloads the octahedral group:

InitGroup[e]; Define[a^2, e]; Define[b^3, e]; Define[c^4, e]
Define[1/a, a]; Define[1/b, b^2]; Define[1/c, c^3]
Define[b.a, a.b.b]; Define[c.a, a.b.c]; Define[c.b, a.c.c]
G = Group[{a, b, c}]

or

```
gap> f:=FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> g:=f/[a^2,b^3,c^4, b*a/(a*b*b), c*a/(a*b*c), c*b/(a*c*c)];;
gap> a:=g.1;; b:=g.2;; c:=g.3;;
gap> G := ListGroup(g);;
```

Non-Interactive Problems

3.8 Prove that the order of Z_n^* is even whenever $n > 2$.
Hint: Find a subgroup of order 2.

3.9 Without using *Mathematica* or GAP, but rather by taking advantage of corollary 3.4, compute 5^{21} (Mod 7) and 7^{21} (Mod 10).

3.10 Show that if H is a subgroup of G, and the left coset xH is also a subgroup of G, then x is in H.

3.11 Show that if an element y of a group G is in the right coset Hx, where H is a subgroup of G, then $Hy = Hx$.

3.12 Let $|G| = 33$. What are the possible orders for the elements of G? Show that G must have an element of order 3.

3.13 Show that proposition 3.2 is still true if n is the product of *three* distinct primes. In fact, many applications of the RSA code use three large primes instead of two.

3.14 Show that proposition 3.2 is no longer true if we let $n = p^2$ for some prime p.

3.15 Show that if G is an abelian group, and X and Y are two subgroups of G, then $X \cdot Y$ is a subgroup of G.

3.16 List all of the left and right cosets of the subgroup { **Stay**, **FlipRt** } of Terry's group. Are the left and right cosets the same?

3.17 List all of the cosets of the subgroup $\{0, 4, 8\}$ of Z_{12}.

3.18 List all of the cosets of the subgroup $\{1, 4\}$ of Z_{15}^*. (See table 1.4.)

For problems **3.19** through **3.21**, write the multiplication table for the following quotient groups:

3.19 $Z_{12}/\{0, 4, 8\}$ **3.20** $Z_{12}/\{0, 6\}$ **3.21** $Z_{15}^*/\{1, 4\}$ (See table 1.4.)

3.22 Find all of the normal subgroups of S_3. (This is Terry's group.)

3.23 Let $\mathbb{Q}$ be the additive group of rational numbers. Show that the group of integers $\mathbb{Z}$ is a normal subgroup of $\mathbb{Q}$. Show that $\mathbb{Q}/\mathbb{Z}$ is an infinite group in which every element has finite order.

3.24 Let G be the group from example 1.4 in section 1.4, the group of *linear functions* of the form $f(x) = mx + b$, with $m, b \in \mathbb{R}$, $m \neq 0$. Let N be the subset of G for which $m = 1$, that is,

$$N = \{\phi(x) = x + b \quad | \quad b \in \mathbb{R}\}.$$

Show that N is a normal subgroup of G. Describe the quotient group G/N.

3.25 Let G be the group of *linear functions* as in problem 3.24. Let T be the subset of G for which $b = 0$, that is,

$$T = \{\phi(x) = mx \quad | \quad m \in \mathbb{R},\, m \neq 0\}.$$

Show that T is a subgroup of G, but not a normal subgroup. If $f(x) = 2x + 3$, describe both the left and right cosets $f \cdot T$ and $T \cdot f$.

3.26 Prove that the quotient group of a cyclic group is cyclic.

3.27 Prove that the quotient group of an abelian group is abelian.

Chapter 4

Mappings between Groups

4.1 Isomorphisms

The quotient group G/M we saw at the end of the last chapter turned out to be very similar to the group S_3. They are technically distinct, since the names for their elements are totally different. Yet we could find a correlation between the elements of the two groups so that the corresponding multiplication tables would have identical color patterns. Here is one such possible correlation between the two groups:

$$\begin{aligned}
e &\leftrightarrow \{e,\, c^2,\, a\cdot b^2\cdot c,\, a\cdot b^2\cdot c^3\}\\
a &\leftrightarrow \{c,\, a\cdot b^2,\, c^3,\, a\cdot b^2\cdot c^2\}\\
b &\leftrightarrow \{b,\, a\cdot b\cdot c,\, b\cdot c^2,\, a\cdot b\cdot c^3\}\\
a\cdot b &\leftrightarrow \{a,\, a\cdot c^2,\, b^2\cdot c,\, b^2\cdot c^3\}\\
b^2 &\leftrightarrow \{a\cdot c,\, b^2,\, a\cdot c^3,\, b^2.c^2\}\\
a\cdot b^2 &\leftrightarrow \{a\cdot b,\, b\cdot c,\, a\cdot b\cdot c^2,\, b\cdot c^3\}
\end{aligned}$$

Suppose we use this correlation to define a *function* $f(x)$ sending each element of S_3 to an element of G/M. Thus,

$$\begin{aligned}
f(e) &= \{e,\, c^2,\, a\cdot b^2\cdot c,\, a\cdot b^2\cdot c^3\}\\
f(a) &= \{c,\, a\cdot b^2,\, c^3,\, a\cdot b^2\cdot c^2\}\\
f(b) &= \{b,\, a\cdot b\cdot c,\, b\cdot c^2,\, a\cdot b\cdot c^3\}\\
f(a\cdot b) &= \{a,\, a\cdot c^2,\, b^2\cdot c,\, b^2\cdot c^3\}\\
f(b^2) &= \{a\cdot c,\, b^2,\, a\cdot c^3,\, b^2.c^2\}\\
f(a\cdot b^2) &= \{a\cdot b,\, b\cdot c,\, a\cdot b\cdot c^2,\, b\cdot c^3\}
\end{aligned}$$

The fact that the corresponding multiplication tables have the same color patterns can now be expressed simply by

$$f(x\cdot y) = f(x)\cdot f(y).$$

Also, the function $f(x)$ maps different elements of S_3 to different elements of G/M. That is, $f(x)$ is one-to-one, or *injective.* Finally, every element of G/M

appears as $f(x)$ for some element x. This is expressed by saying that $f(x)$ is onto, or *surjective.* A function that is both one-to-one and onto is called *bijective.*

DEFINITION 4.1 Let G_1 and G_2 be two groups. An *isomorphism* from G_1 to G_2 is a one-to-one function sending elements of G_1 to elements of G_2 such that

$$f(x \cdot y) = f(x) \cdot f(y) \quad \text{for all } x, y \in G_1.$$

If there exists an isomorphism from G_1 to G_2 that is also onto, then we say that G_1 and G_2 are *isomorphic*, denoted by

$$G_1 \approx G_2.$$

For example,

$$S_3 \approx G/M$$

because of the existence of the function $f(x)$, which we saw was both one-to-one and onto.

One of the important yet extremely hard problems in group theory is to find all of the non-isomorphic groups of a given order. Although this is still an unsolved problem, we have the following upper bound for the number of groups.

PROPOSITION 4.1
There are at most $n^{(n^2)}$ non-isomorphic groups of order n.

PROOF If two groups have the same multiplication table, they are isomorphic, so a group is completely determined by its multiplication table. Notice that each element of this table must be one of n elements, and there are n^2 entries in the table. So there are $n^{(n^2)}$ ways of creating such a table. □

Of course, not very many of these tables will actually form a group. In fact, in some cases we can show that there is only one non-isomorphic group of order n.

PROPOSITION 4.2
For n a positive integer, every cyclic group of order n is isomorphic to Z_n.

PROOF Let G be a group of order n, and let g be a generator of G. Then $g^n = e$, and

$$G = \{e = g^0, g^1, g^2, g^3, \ldots, g^{n-1}\}.$$

Define $f : Z_n \to G$ by

$$f(x) = g^x \qquad (0 \le x \le n-1).$$

That is, f will map the elements of Z_n to elements of G. Clearly f is one-to-one and onto, and we would like to show that it is an isomorphism. Suppose x and y satisfy

$$0 \le x, y \le n-1.$$

We let $z = x+y$ (Mod n). Then we can find an m such that $x+y = mn+z$. Now, $f(x+y) = f(z) = g^z$ by the definition of f. Thus,

$$f(x+y) = g^z = g^{(x+y-mn)} = g^x \cdot g^y \cdot (g^n)^{-m} = g^x \cdot g^y = f(x) \cdot f(y).$$

Since f is an isomorphism of Z_n onto G, we have $G \approx Z_n$. □

In particular if p is prime, corollary 3.3 indicates all groups of order p are cyclic. Thus all groups of order p are isomorphic to Z_p.

For example, there is only one group each, up to isomorphism, of sizes 2, 3, 5, and 7, namely Z_2, Z_3, Z_5, and Z_7. Our goal for this section is to find all of the possible groups, up to isomorphism, up to order 8. To help us in this endeavor we have the following lemma.

LEMMA 4.1

Suppose a group G whose order is greater than 2 has all non-identity elements being of order 2. Then G has a subgroup isomorphic to Z_8^*.

PROOF Since the order of G is greater than 2, there are two distinct elements a and b besides the identity element e. Then we have $a^2 = b^2 = e$. Consider the product $a \cdot b$. It can be neither a nor b since this would imply the other was the identity. On the other hand, $a \cdot b = e$ implies

$$a = a \cdot e = a \cdot (b \cdot b) = (a \cdot b) \cdot b = e \cdot b = b.$$

So $a \cdot b$ is not the identity either. So there must be a fourth element in G, which we will call c, such that $a \cdot b = c$. Since all elements of G are of order 2, we have $c^2 = e$.

Finally, note that

$$b \cdot a = e \cdot b \cdot a \cdot e = a \cdot a \cdot b \cdot a \cdot b \cdot b = a \cdot (a \cdot b)^2 \cdot b = a \cdot c^2 \cdot b = a \cdot e \cdot b = a \cdot b = c.$$

With this we can quickly find the remaining products involving a, b, and c.

$$c{\cdot}a = b{\cdot}a{\cdot}a = b, \qquad c{\cdot}b = a{\cdot}b{\cdot}b = a, \qquad a{\cdot}c = a{\cdot}a{\cdot}b = b, \qquad b{\cdot}c = b{\cdot}b{\cdot}a = a.$$

Hence, the set $H = \{e, a, b, c\}$ is closed under multiplication, contains the identity, and also contains the inverses of every element in the set. Hence, H is a subgroup of G. The multiplication table for H

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

shows that this is isomorphic to Z_8^* using the mapping

$$\begin{aligned} f(e) &= 1, \\ f(a) &= 3, \\ f(b) &= 5, \\ f(c) &= 7. \end{aligned}$$

□

We can now use GAP or *Mathematica*® to find all non-isomorphic groups of order up to 8. For example, if we have a group of order 6, any element of order 6 would imply that it is isomorphic to Z_6. We can't have all non-identity elements to have order 2, or else lemma 4.1 would give a subset of order 4, violating Lagrange's theorem (3.1). Thus, there must be an element b of order 3. Then $N = \{e, b, b^2\}$ is a normal subgroup of order 3 by proposition 3.5. If a^2 is b or b^2, then a is of order 6, so to get something different a^2 must be e. Then since N is normal $b \cdot a$ is either b, $a \cdot b$, or $a \cdot b^2$. GAP can eliminate the first two possibilities:

```
gap> f:=FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g:=f/[a^2, b^3, b*a/b];; a := g.1;; b := g.2;;
gap> Size(g);
3
gap> f:=FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g:=f/[a^2, b^3, b*a/(a*b)];; a := g.1;; b := g.2;;
gap> Order(a*b);
6
```

The first case shows that we no longer have a group with six elements (a becomes e), and the second case still has an element of order 6. The last case of course is the S_3 we are familiar with. Hence, there are two non-isomorphic groups of order 6, Z_6 and S_3.

A similar exhaustive search can be used to find all groups of order 8. If such a group has all non-identity elements of order 2, then by lemma 4.1 there is a subgroup $\{e, a, b, a \cdot b\}$. By problem 1.22, the group is commutative, so if we pick c to be any other element, then $c^2 = e$, $c \cdot a = a \cdot c$, and $c \cdot b = b \cdot c$.

```
gap> f:=FreeGroup("a","b","c");; a:= f.1;; b:=f.2;; c:=f.3;;
gap> g:=f/[a^2, b^2, c^2, b*a/(a*b), c*a/(a*c), c*b/(b*c)];;
gap> Size(g);
8
```

So there is only one group of order 8 for which all non-identity elements are of order 2. But we can find such a group—Z_{24}^*, whose table is given in table 4.1.

TABLE 4.1: Multiplication table for Z_{24}^*

$\cdot$	1	5	7	11	13	17	19	23
1	1	5	7	11	13	17	19	23
5	5	1	11	7	17	13	23	19
7	7	11	1	5	19	23	13	17
11	11	7	5	1	23	19	17	13
13	13	17	19	23	1	5	7	11
17	17	13	23	19	5	1	11	7
19	19	23	13	17	7	11	1	5
23	23	19	17	13	11	7	5	1

If $|G| = 8$ and G is not isomorphic to either Z_8 or Z_{24}^*, then there must be an element b of order 4. Then $S = \{e, b, b^2, b^3\}$ is a normal subgroup, and we can let a be any element not in S. Since G/S has order 2, a^2 must be in S, but if either $a^2 = b$ or $a^2 = b^3$, then a will have order 8. Also, $b \cdot a \notin S$, but $b \cdot a \neq a$, since this would force $b = e$. So a^2 is either e or b^2, and $b \cdot a$ is either $a \cdot b, a \cdot b^2$, or $a \cdot b^3$. These six possibilities can be tried out in GAP or *Mathematica*.

```
gap> f:=FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g:=f/[a^2, b^4, b*a/(a*b^3)];; a := g.1;; b := g.2;;
gap> Size(g);
8
gap> f:=FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g:=f/[a^2/(b^2), b^4, b*a/(a*b^3)];; a := g.1;; b := g.2;;
gap> Size(g);
8
gap> f:=FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g:=f/[a^2, b^4, b*a/(a*b^2)];; a := g.1;; b := g.2;;
gap> Size(g);
2
gap> f:=FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g:=f/[a^2/(b^2), b^4, b*a/(a*b^2)];; a := g.1;; b := g.2;;
gap> Size(g);
2
gap> f:=FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g:=f/[a^2, b^4, b*a/(a*b)];; a := g.1;; b := g.2;;
gap> Size(g);
8
gap> f:=FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g:=f/[a^2/(b^2), b^4, b*a/(a*b)];; a := g.1;; b := g.2;;
gap> Size(g);
8
```

The first possibility gives rise to the group D_4, the symmetry group of the square studied in problem 1.6. The multiplication table shown in table 4.2 can be generated by the *Mathematica* commands:

InitGroup[e];
Define[a^2, e]
Define[b^4, e]
Define[b.a, a.b.b.b]
D4 = Group[{a, b}]

TABLE 4.2: Multiplication table for D_4

$\cdot$	e	a	b	$a \cdot b$	b^2	$a \cdot b^2$	b^3	$a \cdot b^3$
e	e	a	b	$a \cdot b$	b^2	$a \cdot b^2$	b^3	$a \cdot b^3$
a	a	e	$a \cdot b$	b	$a \cdot b^2$	b^2	$a \cdot b^3$	b^3
b	b	$a \cdot b^3$	b^2	a	b^3	$a \cdot b$	e	$a \cdot b^2$
$a \cdot b$	$a \cdot b$	b^3	$a \cdot b^2$	e	$a \cdot b^3$	b	a	b^2
b^2	b^2	$a \cdot b^2$	b^3	$a \cdot b^3$	e	a	b	$a \cdot b$
$a \cdot b^2$	$a \cdot b^2$	b^2	$a \cdot b^3$	b^3	a	e	$a \cdot b$	b
b^3	b^3	$a \cdot b$	e	$a \cdot b^2$	b	$a \cdot b^3$	b^2	a
$a \cdot b^3$	$a \cdot b^3$	b	a	b^2	$a \cdot b$	b^3	$a \cdot b^2$	e

The second possibility produces a new group called the quaternion group Q, described by the following:

InitGroup[e];
Define[a^4, e]
Define[b^2, a^2]
Define[b.a, a.a.a.b]
Q = Group[{a, b}]

Although the group can be defined in terms of only two generators, it is more natural to use the notation that appears in table 4.3.

TABLE 4.3: Multiplication table for Q

$\cdot$	1	I	J	K	-1	$-I$	$-J$	$-K$
1	1	I	J	K	-1	$-I$	$-J$	$-K$
I	I	-1	K	$-J$	$-I$	1	$-K$	J
J	J	$-K$	-1	I	$-J$	K	1	$-I$
K	K	J	$-I$	-1	$-K$	$-J$	I	1
-1	-1	$-I$	$-J$	$-K$	1	I	J	K
$-I$	$-I$	1	$-K$	J	I	-1	K	$-J$
$-J$	$-J$	K	1	$-I$	J	$-K$	-1	I
$-K$	$-K$	$-J$	I	1	K	J	$-I$	-1

The next two possibilities failed to produce a group of order 8, and the last two possibilities are both isomorphic to Z_{15}^* that we have seen before. In summary, we have the following groups up to order 8:

$n = 1$: The one element must be the identity, so we have just the trivial group, $\{e\}$.

$n = 2$: Since 2 is prime, the only non-isomorphic group is Z_2.

$n = 3$: Since 3 is prime, the only non-isomorphic group is Z_3.

$n = 4$: By lemma 4.1, there are two non-isomorphic groups: Z_4 and Z_8^*.

$n = 5$: Since 5 is prime, the only non-isomorphic group is Z_5.

$n = 6$: There are two non-isomorphic groups: Z_6 and the non-abelian group S_3.

$n = 7$: Since 7 is prime, the only non-isomorphic group is Z_7.

$n = 8$: There are three abelian groups, Z_8, Z_{15}^*, and Z_{24}^* and two non-abelian groups, D_4 and Q.

Finally, table 4.4 gives of the number of non-isomorphic groups of order n, when n is not prime.

TABLE 4.4: Groups of order n

n	groups	n	groups	n	groups	n	groups	n	groups
4	2	26	2	46	2	65	1	85	1
6	2	27	5	48	52	66	4	86	2
8	5	28	4	49	2	68	5	87	1
9	2	30	4	50	5	69	1	88	12
10	2	32	51	51	1	70	4	90	10
12	5	33	1	52	5	72	50	91	1
14	2	34	2	54	15	74	2	92	4
15	1	35	1	55	2	75	3	93	2
16	14	36	14	56	13	76	4	94	2
18	5	38	2	57	2	77	1	95	1
20	5	39	2	58	2	78	6	96	230
21	2	40	14	60	13	80	52	98	5
22	2	42	6	62	2	81	15	99	2
24	15	44	4	63	4	82	2	100	16
25	2	45	2	64	267	84	15	102	4

4.2 Homomorphisms

It is easy to see the application of isomorphisms, since these functions show how two groups are essentially the same. But suppose we have a function between two groups for which $f(x \cdot y) = f(x) \cdot f(y)$, but this function may not be one-to-one or onto. Can we still glean some information about the groups from this function?

DEFINITION 4.2 Let G and M be two groups. A function

$$f : G \to M$$

mapping elements of G to elements of M is called a *homomorphism* if it satisfies

$$f(x \cdot y) = f(x) \cdot f(y) \quad \text{for all } x, y \in G.$$

The group G is called the *domain* of the homomorphism, and the group M is called the *target* of the homomorphism. Note that a homomorphism need not be either one-to-one or onto.

Let us look at some examples of homomorphisms.

Example 4.1
Let G be any group, and let M be a group with identity e. If we let

$$f(x) = e \quad \text{for all } x \in G$$

then f will obviously be a homomorphism. This is called the *trivial homomorphism.* □

Example 4.2
Let $\mathbb{R}^* = \mathbb{R} - \{0\}$ be the group of nonzero real numbers under multiplication, and let $f(x) = x^2$. This forms a homomorphism

$$f : \mathbb{R}^* \to \mathbb{R}^*,$$

so homomorphism maps a group onto itself. Note that this homomorphism is neither one-to-one nor onto since $f(-2) = f(2) = 4$, yet there is no real number such that $f(x) = -1$. □

Example 4.3
We can generalize example 4.2 as follows: Let G be any commutative group, and let n be any integer. We can define $f(x) = x^n$. Then $f(x)$ is a homomorphism from G to itself, since

$$f(x \cdot y) = (x \cdot y)^n = x^n \cdot y^n = f(x) \cdot f(y).$$

□

We can prove a few properties that must be true of all homomorphisms.

PROPOSITION 4.3
Let $f : G \to M$ be a homomorphism. Let e denote the identity of G. Then $f(e)$ is the identity element of M.

PROOF Since $e \cdot e = e$ in the group G, we have

$$f(e) = f(e \cdot e) = f(e) \cdot f(e).$$

Multiplying both sides by $[f(e)]^{-1}$ gives us that $f(e)$ is the identity element of M. □

PROPOSITION 4.4
If $f : G \to M$ is a homomorphism, then $f(a^{-1}) = [f(a)]^{-1}$.

PROOF We merely need to show that $f(a) \cdot f(a^{-1})$ is the identity element of M. If e represents the identity element of G, then

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

By proposition 4.3 this is the identity element of M. So

$$f(a^{-1}) = [f(a)]^{-1}.$$ □

To define homomorphisms using *Mathematica* or GAP, we must first define the two groups G and M simultaneously. Let us first load the octahedral group with the following commands:

```
InitGroup[e];
Define[a^2, e]; Define[b^3, e]; Define[c^4, e]
Define[1/a, a]; Define[1/b, b^2]; Define[1/c, c^3]
Define[b.a, a.b.b]; Define[c.a, a.b.c]; Define[c.b, a.c.c]
Oct = Group[{a, b, c}]
```

Next let us define the quaternion group Q from the last section. We will use the letters i and j for the generators.

```
Define[i^4, e]; Define[j^2, i^2]
Define[j.i, i.i.i.j]
Define[1/i, i^3]; Define[1/j, i.i.j]
Q = Group[{i, j}]
```

Notice that we did not perform an **InitGroup** in defining the second group, since this command would have cleared the first group.

We can define the same two groups in GAP as follows:

```
gap> f:=FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> Oct:=f/[a^2,b^3,c^4, b*a*b*a, c*a/(a*b*c), c*b/(a*c*c)];;
gap> SetReducedMultiplication(Oct);
gap> a:=Oct.1;; b:=Oct.2;; c:=Oct.3;;
gap> f:=FreeGroup("i","j");; i:=f.1;; j:=f.2;;
gap> Q:=f/[i^4, i^2*j*2, i*j*i*j];; i:=Q.1;; j:=Q.2;;
```

To define a homomorphism, we only need to tell GAP or *Mathematica* where the generators are sent. Thus, to define the function

$$\begin{aligned} e &\to e, \\ i &\to c^2, \\ i^2 &\to e, \\ i^3 &\to c^2, \\ j &\to a \cdot b^2 \cdot c, \\ i \cdot j &\to a \cdot b^2 \cdot c^3, \\ i^2 \cdot j &\to a \cdot b^2 \cdot c, \\ i^3 \cdot j &\to a \cdot b^2 \cdot c^3; \end{aligned}$$

we have only to define $F[i]$ and $F[j]$. In GAP, this is done with the command

```
gap> F := GroupHomomorphismByImages(Q,Oct,[i,j],[c^2,a*b^2*c]);
[ i, j ] -> [ c^-2, a^-1*b^-1*c ]
```

To plug a value into this function in GAP, we use the `Image` command

```
gap> Image(F,i*j);
a^-1*b^-1*c^-1
```

We can use the `List` with the function feature to see where each element is mapped.

```
gap> List(Q);
[ <identity ...>, i, j, i^2, i*j, i^3, i^2*j, i^3*j ]
gap> List(Q, x->Image(F,x));
[ <identity ...>, c^-2, a^-1*b^-1*c, <identity ...>,
  a^-1*b^-1*c^-1, c^-2, a^-1*b^-1*c, a^-1*b^-1*c^-1 ]
```

To define this homomorphism in *Mathematica*, we have to first explain that **F** will be a homomorphism,

Homomorph[F]

and then define this function on the generators of Q,

Define[F[i], c.c]
Define[F[j], a.b.b.c]

Mathematica can check whether this function is a homomorphism by the command

CheckHomo[F, Q]
True

where Q is the domain of the homomorphism F. Since *Mathematica* returns a value of "True," the function F is indeed a homomorphism. (GAP automatically does this check for you. Had this not been a homomorphism, GAP would have returned "fail.") The command

GraphHomo[F, Q]

will have *Mathematica* draw a picture of this homomorphism as shown in figure 4.1.

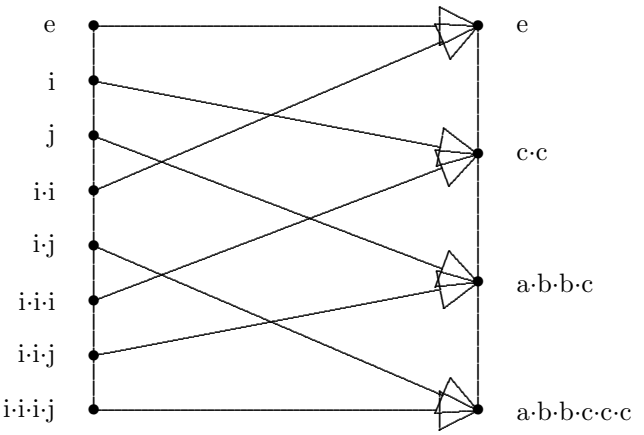


FIGURE 4.1: Diagram of the homomorphism F

We can apply a homomorphism f to a *set* of elements by applying the homomorphism to each element in the set, and consider the set of all possible results. For example, consider the set of real numbers $S = \{-2, -1, 1, 2, 3, 4\}$. Let $f(x)$ be the homomorphism in example 4.2 above, $f(x) = x^2$. Then

$$f(S) = \{1, 4, 9, 16\}.$$

The set $f(S)$ is smaller than the set S, since the homomorphism mapped two elements to both 1 and 4.

To apply the homomorphism to a set of elements in *Mathematica*, we must enclose the set inside an additional pair of curly braces to let *Mathematica* know that we are considering a *set*, rather than a *coset.* For example, the set

S = {i, i.j, i.i.i, i.i.i.j}

is a subset of Q, so we can consider applying F to this set. This is done not by entering **F[S]**, but by the command

F[{S}]

to keep *Mathematica* from interpreting S as a coset of a subgroup, which S happens to be. GAP does not need any extra set of braces.

```
gap> Image(F, [i,i*j, i^3, i^3*j]);
[ c^-2, a^-1*b^-1*c^-1 ]
```

PROPOSITION 4.5
If $f : G \to M$ is a homomorphism and H is a subgroup of G, then $f(H)$ is a subgroup of M.

PROOF We want to show that $f(H)$ is a subgroup using proposition 2.2. If u and v are elements in $f(H)$, there must be elements x and y in H such that $f(x) = u$, and $f(y) = v$.

Then $x \cdot y^{-1}$ is in H, and so

$$f(x \cdot y^{-1}) = f(x) \cdot f(y^{-1}) = f(x) \cdot [f(y)]^{-1} = u \cdot v^{-1}$$

is in $f(H)$. So by proposition 2.2, $f(H)$ is a subgroup of M. □

DEFINITION 4.3 If

$$f : G \to M$$

is a homomorphism, then the group $f(G)$ is called the *range,* or *image* of the homomorphism f. We denote this set by

$$\mathrm{Im}(f).$$

We can also consider taking the inverse homomorphism f^{-1} of an element or a set of elements. Because homomorphisms are not always one-to-one, $f^{-1}(x)$ may not represent a single element. Thus, we will define $f^{-1}(x)$ to be the *set* of numbers such that $f(y) = x$. Likewise, we define

$$f^{-1}(H) = \{y \quad | \quad f(y) \in H\}.$$

We can use *Mathematica*'s **HomoInverse** command to take the inverse homomorphism of an element or set of elements.

HomoInverse[F, c.c, Q]

finds $F^{-1}(c^2)$, using Q is the domain of F. The command

HomoInverse[F, {a, b, a.b.b.c}, Q]

finds the inverse of a set of elements. The corresponding GAP command is `PreImage`:

```
gap> PreImage(F, [c^2, a*b^2*c] );
[ i, j, i^3, i^2*j ]
```

In *Mathematica*, not all of the elements in the set have to be in the image of F, but in GAP they do.

DEFINITION 4.4 If f is a homomorphism from G to M and e is the identity element of M, then we define the *kernel* of f to be the set

$$\mathrm{Ker}(f) = f^{-1}(e).$$

The commands

Kernel[F, Q]

or

```
gap> Kernel(F);
Group([ i^-2 ])
gap> List(last);
[ <identity ...>, i^-2 ]
```

can be used to find the kernel of a homomorphism.

PROPOSITION 4.6
If f is a homomorphism from G to M, then the kernel of f is a normal subgroup of the domain G.

PROOF First we need to show that the kernel of f is a subgroup of G. If e is the identity element of M, and if a and b are two elements of $\mathrm{Ker}(f)$, then

$$f(a \cdot b^{-1}) = f(a) \cdot f(b)^{-1} = e \cdot e^{-1} = e,$$

so $a \cdot b^{-1}$ is also in the kernel of f. Thus, by proposition 2.2, $\mathrm{Ker}(f)$ is a subgroup.

Now let us show that $\mathrm{Ker}(f)$ is a normal subgroup of G. Let a be an element in $\mathrm{Ker}(f)$, and g be any element in G. Then by proposition 3.4, since

$$f(g \cdot a \cdot g^{-1}) = f(g) \cdot f(a) \cdot f(g^{-1}) = f(g) \cdot e \cdot [f(g)]^{-1} = e,$$

$g \cdot a \cdot g^{-1}$ is in $\mathrm{Ker}(f)$, and so $\mathrm{Ker}(f)$ is a normal subgroup. □

Figure 4.1 is very suggestive. The inverse image of any element is a coset of $\{e, i^2\}$. The next proposition explains why this is so.

PROPOSITION 4.7
Let f be a homomorphism from the group G to the group M. Suppose that y is in the image of f, and that $f(x) = y$. Then

$$f^{-1}(y) = x \cdot \text{Ker}(f).$$

PROOF First let us consider an element $z \in x \cdot \text{Ker}(f)$. Then $z = x \cdot k$ for some element k in the kernel of f. Therefore,

$$f(z) = f(x \cdot k) = f(x) \cdot f(k) = f(x) \cdot e = f(x)$$

since k is in $\text{Ker}(f)$. Here, e is the identity element of M. But $f(x) = y$, and so $z \in f^{-1}(y)$. Thus we have proved that

$$f^{-1}(y) \subseteq x \cdot \text{Ker}(f).$$

To prove the inclusion the other way, note that if $z \in f^{-1}(y)$, then $f(z) = y$, and so we have

$$f(x^{-1} \cdot z) = [f(x)]^{-1} \cdot f(z) = y^{-1} \cdot y = e$$

Thus, $x^{-1} \cdot z$ is in the kernel of f, and since $z = x \cdot (x^{-1} \cdot z) \in x \cdot \text{Ker}(f)$, we have

$$x \cdot \text{Ker}(f) \subseteq f^{-1}(y). \qquad \square$$

We now have a quick way to determine if a homomorphism is an isomorphism.

COROLLARY 4.1
Let $f : G \to M$ be a homomorphism. Then f is an injection (one-to-one) if, and only if, the kernel of f is the identity element of G.

PROOF If f is an injection, it is clear that the kernel would just be the identity element. Suppose that the kernel is just the identity. Then proposition 4.7 states that if h is in the image of f, then $f^{-1}(h)$ consists of exactly one element. Therefore, f is one-to-one. $\square$

In particular, if the image of a homomorphism $f : G \to M$ is all of M, and the kernel is $\{e\}$, then $G \approx M$.

We can also consider what happens if we take the inverse image of a subgroup.

COROLLARY 4.2
Let $f : G \to M$ be a homomorphism. Let H be a subgroup of M. Then $f^{-1}(H)$ is a subgroup of G. Furthermore, if H is a normal subgroup of M, then $f^{-1}(H)$ is a normal subgroup of G.

PROOF Let x and y be in $f^{-1}(H)$. Then since $f(x \cdot y^{-1}) = f(x) \cdot [f(y)]^{-1}$, which is in H, we have that $x \cdot y^{-1}$ is in $f^{-1}(H)$. Thus, by proposition 2.2, $f^{-1}(H)$ is a subgroup of G.

Now suppose that H is a normal subgroup of M. Then if y is in $f^{-1}(H)$, and x is in G, then $f(x \cdot y \cdot x^{-1}) = f(x) \cdot f(y) \cdot [f(x)]^{-1}$. Since $f(y)$ is in H, which is normal in M, we have that $f(x) \cdot f(y) \cdot [f(x)]^{-1}$ is in H. Thus, $x \cdot y \cdot x^{-1}$ is in $f^{-1}(H)$, and so by proposition 3.4, $f^{-1}(H)$ is normal in G. □

We are now in a position to show how homomorphisms can be used to reveal relationships between different groups. There are three such relationships to be revealed, and these are covered in the next section.

4.3 The Three Isomorphism Theorems

We have seen in the last section that the kernel K of a homomorphism is always a normal subgroup of the domain G. Furthermore, proposition 4.7 proves what is suggested by figure 4.1, that the inverse image of any element is essentially a coset of K. Hence, the inverse image $f^{-1}(y)$ can be considered as an element of the quotient group G/K. This leads us to the first of three very useful theorems for finding isomorphisms between groups.

THEOREM 4.1: The First Isomorphism Theorem
Let $f : G \to M$ be a homomorphism with $\mathrm{Ker}(f) = K$, *and* $\mathrm{Im}(f) = I$. *Then there is a natural isomorphism*

$$\phi : I \to G/K$$

which is surjective. Thus, $I \approx G/K$.

PROOF It should be noted that this theorem states more than just $I \approx G/K$, but that there is a natural isomorphism between these two groups. This isomorphism is given by

$$\phi(h) = f^{-1}(h).$$

Proposition 4.7 states that whenever h is in the image of f, $f^{-1}(h)$ is a member of the quotient group $G/\mathrm{Ker}(f)$. Thus, $\phi : I \to G/K$ is properly defined.

Let us show that the mapping ϕ is one-to-one. Suppose $\phi(x) = \phi(y)$ for two different elements of I. Then $f(\phi(x)) = f(\phi(y))$. But $f(\phi(x)) = f(f^{-1}(x))$ is the set containing just the element x, and also $f(\phi(y))$ is the set containing just the element y. Thus, $x = y$, and we have shown that ϕ is one-to-one.

Now let us show that ϕ is onto. If xK is an element of G/K, then $f(x) \in I$. Thus,

$$x \in f^{-1}(f(x)) = \phi(f(x)) \in G/K.$$

So we have that x is an element of both cosets xK and $\phi(f(x))$. Since two different cosets have no elements in common, we must have $\phi(f(x)) = xK$. We have therefore that any coset in G/K is mapped by ϕ from an element in I, so ϕ is surjective.

Finally, we want to show that ϕ is a homomorphism. That is, we wish to show that

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

Let $x \in f^{-1}(v)$ and $y \in f^{-1}(w)$. Then $f(x) = v$ and $f(y) = w$, so we have

$$f(x \cdot y) = f(x) \cdot f(y) = v \cdot w.$$

Hence,

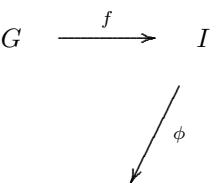
$$x \cdot y \in f^{-1}(v \cdot w).$$

Since $f^{-1}(v) \cdot f^{-1}(w)$ and $f^{-1}(v \cdot w)$ are two cosets in G/K, and both contain the element $x \cdot y$, they must be the same coset. So we have that

$$\phi(v) \cdot \phi(w) = \phi(v \cdot w).$$

□

The natural isomorphism ϕ can be pictured by drawing a diagram:



This diagram suggests that there ought to be a mapping that goes directly from G to $G/\text{Ker}(f)$ without involving the homomorphism f.

PROPOSITION 4.8

Let G be a group, and N be a normal subgroup of G. Then there is a natural isomorphism

$$i_N : G \to G/N$$

given by $i_N(a) = a \cdot N$. This homomorphism is surjective, and $\text{Ker}(i_N) = N$.

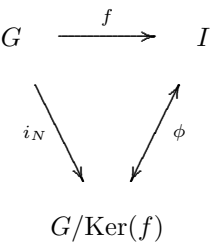


FIGURE 4.2: Commuting diagram for first isomorphism theorem

PROOF To show that i_N is a homomorphism, we note that if a and b are elements of G, then

$$i_N(a \cdot b) = a \cdot b \cdot N = a \cdot N \cdot b \cdot N = i_N(a) \cdot i_N(b).$$

Also, i_N is clearly surjective. To find the kernel of i_N, we note that the identity element of G/N is $eN = N$, and so x is in the kernel if, and only if,

$$i_N(x) = N \Longleftrightarrow x \cdot N = N \Longleftrightarrow x \in N.$$

Therefore, the kernel of i_N is N. □

We call the homomorphism i_N the *canonical homomorphism associated with N*. We can add this homomorphism to our diagram to produce figure 4.2.

The mapping ϕ is shown with a double arrow to show that ϕ is an isomorphism, hence invertible. In this diagram, the functions defined by two paths with the same beginning and ending point produce the same composition function. That is, $\phi(f(x)) = i_N(x)$ and $\phi^{-1}(i_N(x)) = f(x)$. We say that the *diagram is commutative.*

If we consider a group with two normal subgroups, one of which is a subgroup of the other, we begin to see more patterns. Let us reload the octahedral group in GAP, and look at two normal subgroups.

```
gap> f:=FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> Oct:=f/[a^2,b^3,c^4, b*a*b*a, c*a/(a*b*c), c*b/(a*c*c)];;
gap> a:=Oct.1;; b:=Oct.2;; c:=Oct.3;;
gap> G := ListGroup(Oct);
[ <identity ...>, a, b, a*b, b^2, a*b^2, c, a*c, b*c, a*b*c,
  b^2*c, a*b^2*c, c^2, a*c^2, b*c^2, a*b*c^2, b^2*c^2,
  a*b^2*c^2, c^3, a*c^3, b*c^3, a*b*c^3, b^2*c^3, a*b^2*c^3 ]
gap> H:=Group(b,c^2);
Group([ b, c^2 ])
gap> N:=Group(c^2, a*b^2*c);
Group([ c^2, a*b^2*c ])
gap> Size(H);
12
gap> Size(N);
4
```

Both H and N are normal subgroups, so we can consider two different quotient groups.

```
gap> Q1 := RtCoset(G,H);
[ [ <identity ...>, b, a*b^2*c, c^2, b^2, a*b*c, b*c^2,
      a*b^2*c^3, a*c, b^2*c^2, a*b*c^3, a*c^3 ],
  [ a*b^2, a, c, a*b^2*c^2, a*b, b^2*c, a*c^2, c^3, b*c,
      a*b*c^2, b^2*c^3, b*c^3 ] ]
gap> Q2 := RtCoset(G,N);
[ [ <identity ...>, a*b^2*c, c^2, a*b^2*c^3 ],
  [ a*b^2, c, a*b^2*c^2, c^3 ], [b, a*b*c, b*c^2, a*b*c^3 ],
  [ a, b^2*c, a*c^2, b^2*c^3 ], [b^2, a*c, b^2*c^2, a*c^3 ],
  [ a*b, b*c, a*b*c^2, b*c^3 ] ]
```

At this point there doesn't seem to be much connection between these. But notice that N is also a subgroup of H. Is this a normal subgroup? To find out let us determine the left and right cosets of H with respect to N.

```
gap> RtCoset(H,N);
[ [ <identity ...>, b*c^2*b^2, c^2, b^2*c^2*b ],
  [ b, b^2*c^2*b^2, b*c^2, c^2*b ],
  [ b^2, c^2*b^2, b^2*c^2, b*c^2*b ] ]
gap> Q3 := LftCoset(H,N);
[ [ <identity ...>, b*c^2*b^2, c^2, b^2*c^2*b ],
  [ b, b^2*c^2*b^2, b*c^2, c^2*b ],
  [ b^2, c^2*b^2, b^2*c^2, b*c^2*b ] ]
```

Since these two are the same, N must be a normal subgroup of H. We can prove this in general.

LEMMA 4.2
Let N be a normal subgroup of G, and suppose that H is a subgroup of G which contains N. Then N is a normal subgroup of H.

PROOF Since N is a group, and is contained in H, N is a subgroup of H. For any x in H, we have that

$$x \cdot N \cdot x^{-1} = N$$

since x is also in G. Therefore, by proposition 3.4, N is a normal subgroup of H. □

Thus, if both H and N are normal subgroups of G, and $N \subseteq H$, then there will be three quotient groups to consider: G/H, G/N, and H/N. But H/N will be a subgroup of G/N. Could this be a normal subgroup? In the case we are looking at, `Q3`$= H/N$ contains half of the elements of `Q2`$= G/N$, so it is normal, giving us a fourth quotient group:

```
gap> Q4 := RtCoset(Q2, Q3);
[ [ [ <identity ...>, a*b^2*c, c^2, a*b^2*c^3 ],
      [ b, a*b*c, b*c^2, a*b*c^3 ],
      [ b^2, a*c, b^2*c^2, a*c^3 ] ],
  [ [ a*b^2, c, a*b^2*c^2, c^3 ], [ a, b^2*c, a*c^2, b^2*c^3 ],
      [ a*b, b*c, a*b*c^2, b*c^3 ] ] ]
```

Before we try to interpret this mess, let us first see why H/N will be a normal subgroup of G/N in general.

LEMMA 4.3
If H and N are normal subgroups of G, and if N is a subgroup of H, then H/N is a normal subgroup of G/N.

PROOF From lemma 4.2, N is a normal subgroup of H. A typical element of G/N is

$$g \cdot N,$$

where g is an element of G. A typical element of H/N is

$$h \cdot N,$$

where h is an element of H. Thus, H/N is contained in G/N, and so H/N is a subgroup of G/N.

To show that H/N is in fact a normal subgroup of G/N, we will use proposition 3.4. That is, we will see if

$$(g \cdot N) \cdot (h \cdot N) \cdot (g \cdot N)^{-1}$$

will always be in H/N. But this simplifies to $(g \cdot h \cdot g^{-1}) \cdot N$, and $g \cdot h \cdot g^{-1}$ is in H since H is a normal subgroup of G. Therefore, $(g \cdot h \cdot g^{-1}) \cdot N$ is in H/N, and hence H/N is a normal subgroup of G/N. □

The "quotient group of quotient groups" `Q4` $= (G/N)/(H/N)$ is a list containing two lists, each of which contains several lists of elements. If this is too many nested lists for you to handle, imagine what would happen if we removed the innermost brackets. This would simplify the output to just a list of two lists, each of which contains 12 elements. But by looking carefully, we can see that we would get *exactly* `Q1`. We can use the canonical homomorphisms as a tool to strip away these inside level brackets.

THEOREM 4.2: The Second Isomorphism Theorem
Let H and N be normal subgroups of G, and let N be a subgroup of H. Then

$$(G/N)/(H/N) \approx G/H.$$

PROOF We will use the example to guide us in finding a mapping from $(G/N)/(H/N)$ to a set of elements in G. We have a canonical mapping from G to G/N, and another canonical mapping from G/N to $(G/N)/(H/N)$. Let us call these mappings ϕ and f, respectively. Thus, we have the following diagram:

$$\begin{array}{ccc} G & \xrightarrow{\phi} & G/N \\ & & \Big\downarrow f \\ & & (G/N)/(H/N) \end{array}$$

For an element x in G, the composition homomorphism $f(\phi(x))$ gives the element of $(G/N)/(H/N)$ which contains x somewhere inside of it. Since f and ϕ are both surjective, the composition $f(\phi(x))$ is surjective. Thus, the inverse of this homomorphism, $\phi^{-1}(f^{-1}(y))$, gives a list of elements of G that are somewhere inside of the element y. This inverse is the mapping that removes the interior brackets. We only need to check that this is in fact a coset of G/H. Let us determine the kernel of the composition homomorphism $f(\phi(x))$.

Note that if x is in G, and e is the identity element of $(G/N)/(H/N)$, then

$$\begin{aligned} x \in \mathrm{Ker}(f \circ \phi) &\iff f(\phi(x)) = e \\ &\iff \phi(x) \in \mathrm{Ker}(f) = H/N \\ &\iff x \in \phi^{-1}(H/N) = H. \end{aligned}$$

Therefore, the kernel of the composition $f(\phi(x))$ is H, and so from the first isomorphism theorem (4.1),

$$(G/N)/(H/N) \approx G/H. \qquad \square$$

$$\begin{array}{ccc} G & \xrightarrow{\phi} & G/N \\ \Big\downarrow i_H & & \Big\downarrow f \\ G/H & \longleftrightarrow & (G/N)/(H/N) \end{array}$$

FIGURE 4.3: Commuting diagram for second isomorphism theorem

We can describe the second isomorphism theorem visually by the diagram in figure 4.3. Since H is the kernel of the composition homomorphism

$$f(\phi) : G \to (G/N)/(H/N)$$

we have by the first isomorphism theorem that this diagram commutes.

We observed in section 3.3 that the product of two subgroups H and K was not necessarily a subgroup. However, it is possible that if one of the groups is normal, then indeed the product $H \cdot K$ would be a subgroup. Let us try it on the octahedral group we already defined.

```
gap> H := Group(c);;
gap> M := Group(a*b^2*c,c^2);;
gap> HM := Mult(G,H,M);
[ <identity ...>, a*b^2, a*b^2*c, c^2, c, a*b^2*c^2, a*b^2*c^3,
    c^3 ]
gap> Size(Group(last));
8
```

Since the group generated by these eight elements has only eight elements, these eight elements are a subgroup. What happens if we try this in the other order?

```
gap> Mult(G,M,H);
[ <identity ...>, a*b^2, a*b^2*c, c^2, c, a*b^2*c^2, a*b^2*c^3,
    c^3 ]
```

We discovered that not only is $H \cdot M$ a subgroup, but also $M \cdot H$ is exactly the same as $H \cdot M$. It is not hard to see the connection between these two facts.

LEMMA 4.4

Suppose H and K are two subgroups of G. Then $H \cdot K$ is a subgroup if, and only if,

$$H \cdot K = K \cdot H.$$

PROOF Let us first suppose that $H \cdot K$ is a subgroup. Let $h \in H$ and $k \in K$.

We wish to show that the element $h \cdot k$ in $H \cdot K$ is also in $K \cdot H$. Since $H \cdot K$ is a subgroup, $(h \cdot k)^{-1}$ is in $H \cdot K$. Thus, $(h \cdot k)^{-1} = x \cdot y$ for some $x \in H$ and $y \in K$. But then, $h \cdot k = (x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$, and $y^{-1} \cdot x^{-1}$ is in $K \cdot H$. Thus,

$$H \cdot K \subseteq K \cdot H.$$

By a similar argument, the inverse of any element in $K \cdot H$ must be in $H \cdot K$, and so $K \cdot H \subseteq H \cdot K$. Therefore, we have $H \cdot K = K \cdot H$.

Now, let us suppose that $H \cdot K = K \cdot H$. We want to show that $H \cdot K$ is a subgroup. Let $h, x \in H$ and $k, y \in K$ so both $h \cdot k$ and $x \cdot y$ are elements

of $H \cdot K$. By proposition 2.2, it is enough to show that $(h \cdot k) \cdot (x \cdot y)^{-1}$ is in $H \cdot K$. But $(k \cdot y^{-1}) \cdot x^{-1}$ is in $K \cdot H = H \cdot K$, and so there must be two elements $u \in H$ and $v \in K$ such that $(k \cdot y^{-1}) \cdot x^{-1} = u \cdot v$. Then we have

$$(h \cdot k) \cdot (x \cdot y)^{-1} = h \cdot k \cdot y^{-1} \cdot x^{-1} = (h \cdot u) \cdot v$$

which is in $H \cdot K$. Thus, $H \cdot K$ is a subgroup if, and only if, $H \cdot K = K \cdot H$. □

We are now in a position to show that $H \cdot K$ is a subgroup if one of the subgroups H or K is normal.

LEMMA 4.5
If H is a subgroup of G, and N is a normal subgroup of G, then $H \cdot N$ is a subgroup of G.

PROOF If $h \in H$ and $n \in N$, then $h \cdot n \cdot h^{-1}$ is in N, since N is normal. Then

$$h \cdot n = (h \cdot n \cdot h^{-1}) \cdot h$$

is in $N \cdot H$. Thus, $H \cdot N \subseteq N \cdot H$.

By a similar argument $N \cdot H \subseteq H \cdot N$, so $H \cdot N = N \cdot H$. Therefore, $H \cdot N$ is a group by lemma 4.4. □

Lemma 4.5 gives us a second way of forming a new subgroup from two subgroups. The first was given in proposition 2.3—the intersection of two subgroups is again a subgroup. Recall that the *Mathematica* command

Intersection[H, M]

or the GAP function

```
gap> Intersection(H, M);
Group(<fp, no generators known>)
gap> J := List(last);
[ <identity ...>, c^-2 ]
```

finds the intersection of two subgroups. If, as in lemma 4.5, one of the two subgroups is normal, we have the following.

LEMMA 4.6
If N is a normal subgroup of G, and H is a subgroup of G, then

$$H \cap N$$

is a normal subgroup of H.

PROOF Given elements $h \in H$ and $x \in H \cap N$, we note that since x is in N which is a normal subgroup of G, $h \cdot x \cdot h^{-1}$ is in N. Also, x is in H, so $h \cdot x \cdot h^{-1}$ is in H. Thus,

$$h \cdot x \cdot h^{-1} \in H \cap N,$$

and so by proposition 3.4, the intersection is a normal subgroup of H. □

We can ask whether there is a relationship between to two quotient groups $H/(H \cap N)$ and $(H \cdot N)/N$.

```
gap> RtCoset(H,J);
[ [ <identity ...>, c^2 ], [ c, c^3 ] ]
gap> RtCoset(HM,M);
[ [ <identity ...>, a*b^2*c, c^2, a*b^2*c^3 ],
  [ a*b^2, c, a*b^2*c^2, c^3 ] ]
```

Notice that each coset in $H \cdot M/M$ contains one of the cosets from H/J. In fact, if we threw out all elements in a coset of $H \cdot M/M$ that were not an element of H, we would get a coset of H/J. This provides us the mechanism to prove the isomorphism.

THEOREM 4.3: The Third Isomorphism Theorem
Suppose that N is a normal subgroup of G, and that H is a subgroup of G. Then

$$H/(H \cap N) \approx (H \cdot N)/N.$$

PROOF By lemma 4.5, $H \cdot N$ is a subgroup, and by lemma 4.2, N is a normal subgroup of $H \cdot N$. Also, by lemma 4.6, $H \cap N$ is a normal subgroup of H, and so both of the quotient groups are defined.

We will use the two homomorphisms that we discovered were useful for creating a filter, that is,

$$i : H \to H \cdot N$$

$$f : H \cdot N \to (H \cdot N)/N$$

where i is the identity mapping $i(h) = h$, and f is the canonical homomorphism. This gives us the following diagram:

$$\begin{array}{ccc} H & \xrightarrow{\ i\ } & H \cdot N \\ & & \Big\downarrow f \\ & & (H \cdot N)/N \end{array}$$

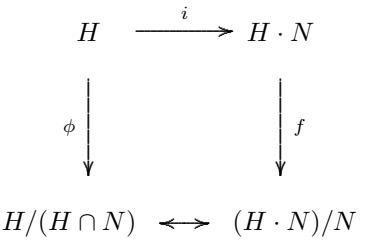


FIGURE 4.4: Commuting diagram for third isomorphism theorem

We can now consider the combination of the two,

$$f(i(h)) : H \to (H \cdot N)/N.$$

We want to find the kernel of this composite homomorphism, for then we can use the first isomorphism theorem (4.1). If we let e denote the identity element of $(H \cdot N)/N$, then

$$\begin{aligned} h \in \text{Ker}(f \cdot i) &\iff f(i(h)) = e \\ &\iff i(h) \in \text{Ker}(f) = N \\ &\iff h \in N \quad \text{and} \quad h \in H \\ &\iff h \in H \cap N. \end{aligned}$$

So by the first isomorphism theorem (4.1), we have

$$(H \cdot N)/N \approx H/(H \cap N). \qquad \square$$

We can describe the third isomorphism theorem (4.3) pictorially through the diagram in figure 4.4, which is commutative according to the first isomorphism theorem (4.1): Note that this diagram demonstrates that

$$|H|/|H \cap N| = |H \cdot N|/|N|.$$

We conclude this chapter by showing that $|H|/|H \cap N| = |H \cdot N|/|N|$ even when neither of the groups H nor N is a normal subgroup.

PROPOSITION 4.9

Let H and K be two subgroups of a finite group G. Then the number of elements in the product $H \cdot K$ is given by

$$|H \cdot K| = \frac{|H|\,|K|}{|H \cap K|}.$$

PROOF Even though $H \cdot K$ is not a group, it still makes sense to consider the set of left cosets $(H \cdot K)/K$. A typical left coset belonging to $(H \cdot K)/K$ would be $h \cdot k \cdot K$, where h is an element of H, and k is an element of K. By lemma 3.1, all cosets contain $|K|$ elements, and by lemma 3.2 two cosets would intersect if, and only if, they are equal. Thus the elements of $H \cdot K$ are distributed into non-overlapping cosets, each having $|K|$ elements. Thus, the number of cosets in $(H \cdot K)/K$ is

$$|(H \cdot K)/K| = \frac{|H \cdot K|}{|K|}.$$

Likewise, we have

$$|H/(H \cap K)| = \frac{|H|}{|H \cap K|}.$$

Thus, if we can show that $|H/(H \cap K)| = |(H \cdot K)/K|$, we will have proven the proposition. Let us define a mapping (not a homomorphism) that will relate the elements of these two sets. Let

$$\phi : (H \cdot K)/K \to H/(H \cap K)$$

be defined by

$$\phi(h \cdot K) = h \cdot (H \cap K).$$

To see that this is well defined, note that if $h \cdot K = x \cdot K$ for two elements h and x in H, then $h^{-1} \cdot x \cdot K = K$, so $h^{-1} \cdot x$ must be in K. Since h and x are also in H, $h^{-1} \cdot x$ is in the intersection, and so

$$x \cdot (H \cap K) = h \cdot (h^{-1} \cdot x) \cdot (H \cap K) = h \cdot (H \cap K).$$

On the other hand, if $h \cdot (H \cap K) = x \cdot (H \cap K)$, then $h^{-1} \cdot x$ would have to be in the intersection of H and K. So then, $h \cdot K = x \cdot K$. Hence the mapping is one-to-one. It is clear that the mapping is also surjective. Hence, ϕ is a bijection, and the proposition is proved. □

Problems for Chapter 4

Interactive Problems

4.1 Prove that there are exactly two non-isomorphic groups of order 10. Find these two groups, and have *Mathematica* or GAP produce the multiplication tables.

Hint: Follow the logic for $n = 6$.

For problems **4.2** through **4.4**: Each of the following groups is of order 8. Which of the known five groups (Z_8, Z_{24}^*, Z_{15}^*, D_4, or Q) is each of these isomorphic to? First have GAP or *Mathematica* display a table of the new group, and then rearrange the elements of one of the five known groups so that the color/number patterns in the two tables are identical.

4.2 Z_{16}^* **4.3** Z_{20}^* **4.4** Z_{30}^*

4.5 Define Terry's group in *Mathematica* with the command

InitTerry

and then define the group S_3 using "Stay" as the identity element.

Define[a^2, Stay]
Define[b^3, Stay]
Define[1/a, a]
Define[1/b, b^2]
Define[b.a, a.b.b]
S3 = Group[{a, b}]

Now define an isomorphism F from S_3 to Terry's group. Use *Mathematica*'s **CheckHomo** command to verify that your function is a homomorphism. Finally, find the kernel of F to prove that F is an isomorphism.

4.6 Use *Mathematica* or GAP to find all of the homomorphisms from S_3 to itself. Label these homomorphisms $F1$, $F2$, $F3$, etc. How many of these are isomorphisms? The following reloads S_3 into *Mathematica*:

InitGroup[e];
Define[a^2, e]; Define[b^3, e]
Define[1/a, a]; Define[1/b, b^2]
Define[b.a, a.b.b]
S3 = Group[{a, b}]

or, to load this group in GAP:

```
gap> f := FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g := f/[a^2, b^3, b*a/(a*b^2)];; a := g.1;; b := g.2;;
gap> S3 := Group(a,b);
Group([ a, b ])
gap> List(S3);
[ <identity ...>, a, b, a*b, a*b*a, b*a ]
```

Non-Interactive Problems

4.7 Prove that if f is a surjective isomorphism from a group G to a group M, then f^{-1} is a surjective isomorphism from M to G.

4.8 If G_1, G_2, and G_3 are three groups, and f is an isomorphism from G_1 to G_2, and ϕ is an isomorphism from G_2 to G_3, prove that $\phi(f)$ is an isomorphism from G_1 to G_3.

4.9 Find an isomorphism between the group consisting of the four complex numbers

$$\{1, -1, i, -i\}$$

and the group Z_4.

For problems **4.10** through **4.18**: Find an isomorphism between the two groups.

4.10 Z_6 and Z_7^*
4.11 Z_6 and Z_9^*
4.12 Z_6 and Z_{14}^*
4.13 Z_6 and Z_{18}^*
4.14 Z_{10} and Z_{11}^*
4.15 Z_{10} and Z_{22}^*
4.16 Z_{12} and Z_{13}^*
4.17 Z_{12} and Z_{26}^*
4.18 Z_8^* and Z_{12}^*

4.19 Let G be an arbitrary group. Prove or disprove that $f(x) = x^{-1}$ is an isomorphism from G to G.

4.20 Prove that any infinite cyclic group is isomorphic to $\mathbb{Z}$.

4.21 Prove that if *both* H and N are normal subgroups of a group G, then $H \cdot N$ is a *normal* subgroup of G.

4.22 If ϕ is a homomorphism from an abelian group G to a group M, show that $\mathrm{Im}(\phi)$ is abelian.

4.23 If ϕ is a homomorphism from a cyclic group G to a group M, show that $\mathrm{Im}(\phi)$ is a cyclic group.

4.24 Let X, Y, and Z be three subgroups of a finite group G, with Y normal. Use proposition 4.9 to find a formula for the number of elements in $X \cdot Y \cdot Z$.

4.25 Let $\mathbb{Z}$ be the group of integers using addition. Show that the function $\phi(x) = 2x$ is a homomorphism from $\mathbb{Z}$ to itself. What is the image of this homomorphism? What is the kernel?

4.26 Let $\mathbb{Z}$ be the group of integers using addition. Show that the function $\phi(x) = -x$ is a homomorphism from $\mathbb{Z}$ to itself. Show that this mapping is in fact one-to-one and onto.

4.27 Let $\mathbb{Z}$ be the group of integers using addition. Show that the function $\phi(x) = x + 3$ is *not* a homomorphism from $\mathbb{Z}$ to itself.

4.28 Let $\mathbb{R}^*$ denote the group of nonzero real numbers, using multiplication as the operation. Let $\phi(x) = x^6$. Show that ϕ is a homomorphism from $\mathbb{R}^*$ to $\mathbb{R}^*$. What is the kernel of this homomorphism? What is the image of the homomorphism?

4.29 Let $\mathbb{R}^*$ denote the group of nonzero real numbers, using multiplication as the operation. Let $\phi(x) = 2x$. Show that ϕ is *not* a homomorphism from $\mathbb{R}^*$ to $\mathbb{R}^*$.

4.30 Let $\mathbb{R}^*$ denote the group of nonzero real numbers, using multiplication as the operation. Recall that $\mathbb{R}$ is the group of real numbers using addition for the operation. Let $\phi(x) = \ln|x|$. Show that ϕ is a homomorphism from $\mathbb{R}^*$ to $\mathbb{R}$. What is the kernel of this homomorphism?

4.31 Let $\mathbb{R}^*$ denote the group of nonzero real numbers, using multiplication as the operation. Recall that $\mathbb{R}$ is the group of real numbers using addition for the operation. Let $\phi(x) = e^x$. Show that ϕ is a homomorphism from $\mathbb{R}$ to $\mathbb{R}^*$. What is the image of this homomorphism?

4.32 Let $\mathbb{R}[t]$ denote the group of all polynomials in t with real coefficients under addition, and let ϕ denote the mapping $\phi(f) = f'$, which sends each polynomial to its derivative. Show that ϕ is a homomorphism from $\mathbb{R}[t]$ to $\mathbb{R}[t]$. What is the kernel of ϕ?

4.33 Let $\mathbb{R}[t]$ denote the group of all polynomials in t with real coefficients under addition. Prove that the mapping from $\mathbb{R}[t]$ into $\mathbb{R}$ given by $f(t) \to f(3)$ is a homomorphism. Give a description of the kernel of this homomorphism.

4.34 Find a homomorphism ϕ from Z_{15}^* to Z_{15}^* with kernel $\{1, 11\}$ and with $\phi(2) = 7$.

4.35 Find a homomorphism ϕ from Z_{30}^* to Z_{30}^* with kernel $\{1, 11\}$ and with $\phi(7) = 13$.

4.36 Find a homomorphism from the quaternion group Q onto Z_8^*.
Hint: The kernel must be a normal subgroup of order 2. See table 4.3 for a multiplication table of Q.

4.37 Let k be a divisor of n. Show that the mapping $\phi(x) = x \pmod{k}$ is a homomorphism from Z_n^* to Z_k^*. Find a formula for the number of elements in the kernel.

4.38 Find all of the homomorphisms from Z_4 to Z_8^*.

4.39 Find all of the homomorphisms from Z_8^* to S_3.

4.40 Prove that there can be no nontrivial homomorphisms from S_3 to Z_3.
Hint: What are the normal subgroups of S_3?

4.41 Suppose that there is a homomorphism from a finite group G onto Z_6. Prove that there are normal subgroups of G with index 2 and 3.

4.42 Suppose that H and K are distinct subgroups of G of index 2. Prove that $H \cap K$ is a normal subgroup of G of index 4 and that $G/(H \cap K) \approx Z_8^*$.

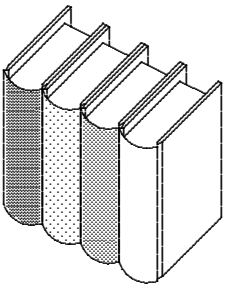
Chapter 5

Permutation Groups

5.1 Symmetric Groups

In this chapter we will explore *permutation groups* or the *symmetric groups,* which have important applications. In fact, we have already seen one example of a symmetric group, S_3. We can easily generalize this group, and consider the group of all permutations of n objects. For example, with four books the beginning position would be

InitBooks[4]



There are six *Mathematica*® operations that rearrange these books.

MoveBooks[First]	swap the first two books.
MoveBooks[Last]	swap the last two books.
MoveBooks[Left]	move the first book to the end, sliding the other books to the left.
MoveBooks[Right]	move the last book to the beginning, sliding the other books to the right.
MoveBooks[Rev]	reverse the order of the books.
MoveBooks[Stay]	leave the books as they are.

For three books, any permutation can be obtained by just one of these six commands. But with four books it is a bit tricky to arrange the books in a particular order. With even more books, it becomes very cumbersome. Thus, we introduce a new notation for a permutation that explicitly states where each book ends up. For example, after a **MoveBooks[Left]** command we find that the 1st book ended up in the 4th position, the 2nd book ended up

in the 1st position, the 3rd book ended in the 2nd position, and the 4th book ended in the 3rd position.

The permutation can be represented writing the ending position under the starting position for the four objects:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}.$$

We can multiply the permutations using the new notation. For example, to calculate **Left·Last**, we have

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}.$$

On the other hand, **Last·Left** is given by

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}.$$

We can interpret each permutation as a *function* whose domain is a subset of the integers. For example, the permutations $f(x) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$ and $\phi(x) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$ can be thought of as two functions for which

$$\begin{aligned} f(1) &= 2 & \phi(1) &= 2 \\ f(2) &= 3 & \phi(2) &= 3 \\ f(3) &= 1 & \phi(3) &= 4 \\ f(4) &= 4 & \phi(4) &= 1. \end{aligned}$$

Note that $f(x)$ appears directly below x in the permutation $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$. The product of the permutations is the same as the composition of the two functions. Thus,

$$\begin{aligned} \phi(f(1)) &= \phi(2) = 3 \\ \phi(f(2)) &= \phi(3) = 4 \\ \phi(f(3)) &= \phi(1) = 2 \\ \phi(f(4)) &= \phi(4) = 1. \end{aligned}$$

Thus, the composition function of doing f first, and then ϕ, is $f \cdot \phi = \phi(f(x)) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}$. Note that permutations are always performed from left to right. However, composition of functions, such as $\phi(f(x))$, is performed from right to left (the inside function is applied first). So when representing a

permutation as a function, we must reverse the order that the functions appear in the formula.

To enter a permutation into *Mathematica* or GAP, only the bottom line is needed. A permutation in S_n can be entered in *Mathematica* as

$$P[x_1, x_2, x_3, \ldots, x_n],$$

where $x_1, x_2, x_3, \ldots x_n$ are distinct integers ranging from 1 to n. This permutation corresponds to the function

$$\begin{aligned} f(1) &= x_1 \\ f(2) &= x_2 \\ f(3) &= x_3 \\ &\cdots \\ f(n) &= x_n. \end{aligned}$$

Thus the *Mathematica* product

P[4,3,5,1,2] . P[5,4,1,2,3]

yields $P[2, 1, 3, 5, 4]$. On the other hand,

P[5,4,1,2,3] . P[4,3,5,1,2]

yields $P[2, 1, 4, 3]$.

Since the composition function maps 5 to itself, *Mathematica* drops the 5, treating this as a permutation on four elements instead.

When we enter the same permutations into GAP, they become *transformations*.

```
gap> P([4,3,5,1,2]);
Transformation( [ 4, 3, 5, 1, 2 ] )
gap> P([4,3,5,1,2])*P([5,4,1,2,3]);
Transformation( [ 2, 1, 3, 5, 4 ] )
gap> P([5,4,1,2,3])*P([4,3,5,1,2]);
Transformation( [ 2, 1, 4, 3, 5 ] )
```

Note that GAP does not drop the final 5 as *Mathematica* did.

Mathematica can use the circle graphs on the set $\{1, 2, \ldots, n\}$ to visualize permutations. For example,

CircleGraph[{1, 2, 3, 4, 5}, P[4, 3, 5, 1, 2]]

produces the circle graph on the left side of figure 5.1. The dotted arrows form a triangle that connects 2, 3, and 5, while the dotted "double arrow" connects 1 and 4. So this circle graph reveals some additional structure to the permutation that we will study later.

We can graph two or more permutations simultaneously. The command

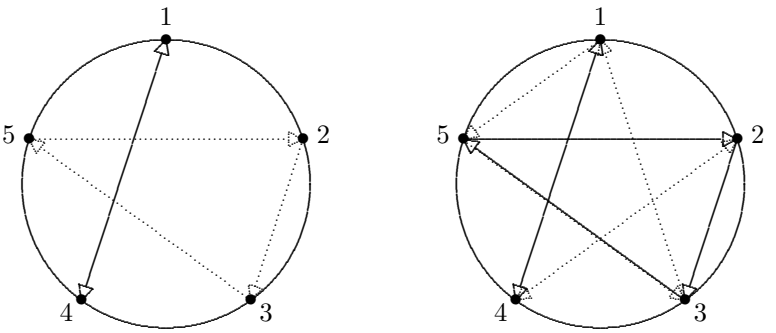


FIGURE 5.1: Circle graphs of permutations

CircleGraph[{1, 2, 3, 4, 5}, P[4, 3, 5, 1, 2], P[5, 4, 1, 2, 3]]

produces the circle graph on the right of figure 5.1. Here, the solid arrows represent the permutation $P[4,3,5,1,2]$, while the dotted arrows represent $P[5,4,1,2,3]$. If one imagines a permutation formed by traveling first through a solid arrow, and then through a dotted arrow, one obtains the permutation $P[2,1,3,5,4]$, which is $P[5,4,1,2,3] \cdot P[4,3,5,1,2]$.

The inverse of a permutation can be found using *Mathematica* or GAP.

P[4,3,5,1,2]^(–1)

```
gap>PermInv(P([4,3,5,1,2]));
Transformation( [ 4, 5, 2, 1, 3 ] )
```

The circle graph of the inverse permutation is similar to the circle graph of $P[4,3,5,1,2]$ except that all arrows are going in the opposite direction. The product of a permutation and its inverse of course will yield the identity element, denoted by $P[\ ]$ in *Mathematica*,

P[4,3,5,1,2] . P[4,5,2,1,3]
$P[\ ]$

or in GAP,

```
gap> P([4,3,5,1,2])*P([4,5,2,1,3]);
Transformation( [ 1, 2, 3, 4, 5 ] )
```

Both *Mathematica* and GAP can treat a permutation as a function, but *Mathematica*'s notation is more standard:

P[4,3,5,1,2][2]

yields $f(2) = 3$. To do the same thing in GAP, we raise 2 to the power of the transformation.

```
gap>2^P([4,3,5,1,2]);
3
```

In spite of the simplicity of the notations for a permutation, we will find that there is yet another notation that is even more concise. We will study this in the next section.

5.2 Cycles

Although GAP is able to multiply transformations together, GAP prefers that permutations be entered in terms of *cycles*. In the circle graph for the permutation $P[4,3,5,1,2]$, we saw that the arrows connecting 2, 3, and 5 were of one color, while a different colored arrow connected 1 and 4. By experimenting, we find that other permutations such as $P[4,5,2,3,1]$ have circle graphs with arrows of only one color, as in figure 5.2.

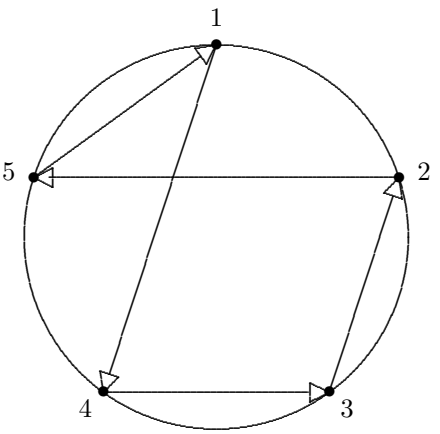


FIGURE 5.2: Circle graph of a cycle

These arrows indicate that the permutation can be expressed by a single chain

$$1 \to 4 \to 3 \to 2 \to 5 \to 1.$$

Other permutations, such as $P[2,4,1,6,5,3]$, have every *straight* arrow of the same color, even though there is one point (5) that maps to itself. We can still express this permutation as a single chain

$$1 \to 2 \to 4 \to 6 \to 3 \to 1,$$

if we stipulate that all numbers that are not mentioned in the chain map to themselves.

DEFINITION 5.1 Any permutation that can be expressed as a single chain is called a *cycle.* A cycle that moves exactly r of the numbers is called an *r-cycle.*

Let us introduce a concise notation for cycles. We can abbreviate a chain such as

$$1 \to 2 \to 4 \to 6 \to 3 \to 1,$$

to simply

$$(1\,2\,4\,6\,3).$$

This is called the *cycle notation* for the permutation. Each number in the cycle is mapped to the next number. The last number in the cycle is mapped to the first number. In general, the r-cycle

$$(i_1 i_2 i_3 \dots i_r)$$

represents the permutation that maps i_1 to i_2 , i_2 to i_3, etc., and finally i_r back to i_1. Notice that

$$(i_1 i_2 i_3 \dots i_r)^{-1} = (i_r i_{r-1} \dots i_3 i_2 i_1),$$

so the inverse of an r-cycle will always be an r-cycle. The identity element can be written as the 0-cycle ( ).

A 1-cycle is really an oxymoron (a contradiction in terms), for there can be no one-to-one mapping that moves only one element and leaves the others fixed. We say that an r-cycle is a *nontrivial r-cycle* if $r > 1$.

Most permutations cannot be written as a single chain. This is evident from looking at the circle graph for the permutation $P[4, 3, 5, 1, 2]$. However, the two different types of arrows suggest that this permutation could be expressed as *two* cycles, one that represents the triangle from 2 to 3 to 5, and back to 2, and the other that exchanges 1 and 4. These two permutations are $P[1, 3, 5, 4, 2]$ and $P[4, 2, 3, 1, 5]$. These two cycles multiply together to give $P[4, 3, 5, 1, 2]$. In fact, this product can be done in either order. If we write these two permutations in cycle notation,

$$P[1, 3, 5, 4, 2] = (2\,3\,5), \qquad P[4, 2, 3, 1, 5] = (1\,4),$$

we notice that there are no numbers in common between these two cycles.

DEFINITION 5.2 Two cycles

$$(i_1 i_2 i_3 \dots i_r) \qquad \text{and} \qquad (j_1 j_2 j_3 \dots j_s)$$

are *disjoint* if none of the i's are equal to any of the j's.

LEMMA 5.1

Let x be an element of S_n which is not the identity. Then x can be written as a product of nontrivial disjoint cycles. This representation of x is unique up to the rearrangement of the cycles.

PROOF Let us say that x fixes the integer i if $x(i) = i$. We will use induction on the number of integers not left fixed by x, denoted by m. Because x is not the identity, there is at least one integer not fixed by x. In fact, m must be at least 2, for the first integer must have somewhere to go.

If $m = 2$, then only two numbers i_1 and i_2 are moved. Since these are the only two integers not fixed, x must be a 2-cycle $(i_1 i_2)$.

We now will assume by induction that the lemma is true whenever the number of integers not left fixed by x is fewer than m. Let i_1 be one integer that is not fixed, and let $i_2 = x(i_1)$. Then $x(i_2)$ cannot be i_2 for x is one-to-one, and if $x(i_2)$ is not i_1, we define $i_3 = x(i_2)$. Likewise, $x(i_3)$ cannot be either i_2 or i_3, since x is one-to-one. If $x(i_3)$ is not i_1, we define $i_4 = x(i_3)$.

Eventually this process must stop, for there are only m elements that are not fixed by x. Thus, there must be some value k such that $x(i_k) = i_1$. Define the permutation y to be the k-cycle $(i_1 i_2 i_3 \dots i_k)$. Then $x \cdot y^{-1}$ fixes all of the integers fixed by x, along with $i_1, i_2, i_3, \dots, i_k$. By induction, since there are fewer integers not fixed by $x \cdot y^{-1}$ then by x, $x \cdot y^{-1}$ can be expressed by a series of nontrivial disjoint cycles $c_1 \cdot c_2 \cdot c_3 \cdots c_t$. Moreover, the integers appearing in $c_1 \cdot c_2 \cdot c_3 \cdots c_t$ are just those that are not fixed by $x \cdot y^{-1}$. Thus, $c_1 \cdot c_2 \cdot c_3 \cdots c_t$ are disjoint from y. Finally, we have

$$x = y \cdot c_1 \cdot c_2 \cdot c_3 \cdots c_t.$$

Therefore, x can be written as a product of disjoint nontrivial cycles. By induction, every permutation besides the identity can be written as a product of nontrivial disjoint cycles.

For the uniqueness, suppose that a permutation x has two ways of being written is terms of nontrivial disjoint cycles:

$$x = c_1 \cdot c_2 \cdot c_3 \cdots c_r = d_1 \cdot d_2 \cdot d_3 \cdots d_s.$$

For any integer i_1 not fixed by x, one and only one cycle must contain i_1. Suppose that cycle is $c_j = (i_1 i_2 i_3 \dots i_q)$. But by the way we constructed the cycles above, this cycle must also be one of the d_k's. Thus, each cycle c_j is equal to d_k for some k. By symmetry, each d_k is equal to c_j for some j. Thus, the two ways of writing x in terms of nontrivial disjoint cycles are merely rearrangements of the cycles. □

Lemma 5.1 gives us a succinct way to express permutations. *Mathematica* uses the notation

C[2,3,4,5]

to denote the cycle (2 3 4 5). *Mathematica* can multiply two cycles together,

C[2,3,4,5] . C[1,2,4]

forming the answer as a product of two disjoint cycles. In GAP, the cycles are expressed using only parentheses. Thus, this product in GAP is written

```
gap> (2,3,4,5)*(1,2,4);
(1,2,3)(4,5)
```

Note that when two cycles are disjoint, we do not need the times sign between them. In fact, GAP sees $(1,2,3)(4,5)$ not as a product, but as a single permutation. We call this the *cycle decomposition* of the permutation. We can convert from the cycle notation to the permutation and vice versa in GAP with the commands

```
gap> CycleToPerm( (1,3,4)(2,5) );
Transformation( [ 3, 5, 4, 1, 2 ] )
gap> PermToCycle(last);
(1,3,4)(2,5)
```

These commands also work in *Mathematica*.

CycleToPerm[C[1,3,4] . C[2,5]]
$P[3,5,4,1,2]$
PermToCycle[P[4,6,1,8,2,5,7,3]]
$C[1,4,8,3] \cdot C[2,6,5]$

We may even mix the two notations in *Mathematica* within an expression, such as:

C[1,2,3] . P[3,1,2,5,4] . C[4,5]

Whenever *Mathematica* encounters a mixture like this, it puts the answer in terms of cycles. In this case the result is the identity permutation, so *Mathematica* returns $C[\ ]$, which corresponds to the 0-cycle ( ).

The group S_4 is generated by $P[2,1]$, $P[2,3,1]$, and $P[4,3,2,1]$. Thus, we can produce the symmetric group S_4 in *Mathematica*.

S4 = Group[{P[2,1], P[2,3,1], P[2,3,4,1]}]

To form a group of permutations in GAP, we *must* use the cycle notations. Thus, S_4 is created in GAP with the command

```
gap> S4 := Group( (1,2), (1,2,3), (1,2,3,4) );
Group([ (1,2), (1,2,3), (1,2,3,4) ])
gap> List(S4);
[ (), (1,4)(2,3), (1,2)(3,4), (1,3)(2,4), (2,4,3), (1,4,2),
  (1,2,3), (1,3,4), (2,3,4), (1,4,3), (1,2,4), (1,3,2), (3,4),
  (1,4,2,3), (1,2), (1,3,2,4), (2,4), (1,4,3,2), (1,2,3,4),
  (1,3), (2,3), (1,4), (1,2,4,3), (1,3,4,2) ]
gap> Size(S4);
24
```

The size of S_4 is 24 elements, since there are 24 ways to arrange four books on a shelf. In general, the size of S_n is $n!$, where

$$n! = n \cdot (n-1) \cdot (n-2) \cdot (n-3) \cdot \ldots \cdot 2 \cdot 1.$$

The number $n!$ is read "n factorial." Table 5.1 gives a short table for $n!$.

TABLE 5.1: $n!$ for $n \leq 10$

1!	= 1	6!	= 720
2!	= 2	7!	= 5040
3!	= 6	8!	= 40320
4!	= 24	9!	= 362880
5!	= 120	10!	= 3628800

Both S_4 and the octahedral group have 24 elements, so we could ask if these two groups are isomorphic. The octahedral group can be reloaded by the commands

InitGroup[e];
Define[a^2, e]; Define[b^3, e]; Define[c^4, e]
Define[1/a, a]; Define[1/b, b^2]; Define[1/c, c^3]
Define[b.a, a.b.b]; Define[c.a, a.b.c]; Define[c.b, a.c.c]
G = Group[{a, b, c}]

or, in GAP,

```
gap> f:=FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> G:=f/[a^2,b^3,c^4,b*a/(a*b*b),c*a/(a*b*c),c*b/(a*c*c)];;
gap> a:=G.1;; b:=G.2;; c:=G.3;;
```

Let us begin by defining a homomorphism from the subgroup generated by a and b to S_3, since we know that this is an isomorphism.

```
gap> H := Group(a,b);
Group([ a, b ])
gap> F := GroupHomomorphismByImages(H,S4,[a,b],[(1,2),(1,2,3)]);
[ a, b ] -> [ (1,2), (1,2,3) ]
```

To define this homomorphism in *Mathematica*, we have to be a bit more careful, since the identity of G is e, whereas the identity of S_4 is ( ). We accomplish this change of identity notation by specifying the identity element of the target group when we define the homomorphism.

Homomorph[F, P[ ]]
Define[F[a], P[2,1]]
Define[F[b], P[2,3,1]]
H = Group[{a, b }];
CheckHomo[F, H]

Both *Mathematica* and GAP show that so far, this homomorphism is consistent. To finish this homomorphism we only need to define $F[c]$. Since c must map to an element of order 4, there are six possibilities: (1234), (1243), (1324), (1342), (1423), and (1432). A little trial and error finds the right combination.

Define[F[c], P[2,4,1,3]]
CheckHomo[F, G]

In GAP, we have to redefine the homomorphism from scratch each time, until we get one that works.

```
gap> F := GroupHomomorphismByImages(G,S4,[a,b,c],
> [(1,2),(1,2,3),(1,2,3,4)]);
fail
gap> F := GroupHomomorphismByImages(G,S4,[a,b,c],
> [(1,2),(1,2,3),(1,2,4,3)]);
[ a, b, c ] -> [ (1,2), (1,2,3), (1,2,4,3) ]
```

Next we want to see that F is an isomorphism by showing that the kernel of F,

```
gap> List(Kernel(F));
[ <identity ...> ]
```

or, in *Mathematica*,

Kernel[F, G]

reveals the kernel is just the identity. Then by the pigeonhole principle, the image of F must be all of S_4, so $G \approx S_4$.

In *Mathematica*, we can create a circle graph of a cycle, or product of cycles, just as we did for permutations. We can even treat a cycle as a function, as we did for permutations. For example,

C[1,4,8,3][3]

determines where the cycle (1483) sends the number 3. However, to evaluate a product of cycles at a given number, an extra pair of parentheses is needed:

(C[1,4,8,3] . C[2,6,5])[5]

In GAP, evaluating a cycle or product of disjoint cycles at a number is accomplished by raising the number to the cycle. Thus,

```
gap> 3^(1,4,8,3);
1
gap> 5^(1,4,8,3)(2,6,5);
2
```

As long as the multiplication sign is not between the disjoint cycles, GAP sees this as a single permutation, so no parentheses are needed.

DEFINITION 5.3 A *transposition* is a 2-cycle $(i_1 i_2)$, where $i_1 \neq i_2$.

Observe that i_1 can be any of the n numbers, and i_2 can be any of the remaining $n-1$ numbers, but this counts each transposition twice, since $(i_1 i_2) = (i_2 i_1)$. Thus, there are

$$\frac{n(n-1)}{2} = \frac{n^2 - n}{2}$$

transpositions of S_n.

LEMMA 5.2
For $n > 1$, the set of transpositions in S_n generates S_n.

PROOF We need to show that every element of S_n can be written as a product of transpositions. The identity element can be written as (12)(12), so we let x be a permutation that is not the identity. By lemma 5.1, we can express x as a product of nontrivial disjoint cycles:

$$x = (i_1 i_2 i_3 \dots i_r) \cdot (j_1 j_2 \dots j_s) \cdot (k_1 k_2 \dots k_t) \cdot \dots .$$

Now, consider the product of transpositions

$$(i_{r-1} i_r) \cdot (i_{r-2} i_{r-1}) \cdots (i_2 i_3) \cdot (i_1 i_2) \cdot (j_{s-1} j_s) \cdots (j_1 j_2) \cdot (k_{t-1} k_t) \cdots (k_1 k_2) \cdot \dots .$$

Note that this product is equal to x. (Recall that we are working from left to right.) Therefore, we have expressed every element of S_n as a product of transpositions. □

Of course, a particular permutation can be expressed as a product of transpositions in more than one way. But an important property of the symmetric groups is that the number of transpositions used to represent a given permutation will always have the same parity, that is, even or odd. To show this, we will first prove the following lemma.

LEMMA 5.3
The product of an odd number of transpositions in S_n cannot equal the identity element.

PROOF Since S_2 only contains one transposition, (1 2), raising this to an odd power will not be the identity element, so the lemma is true for the

case $n = 2$. So by induction we can assume that the lemma is true for S_{n-1}. Suppose that there is an odd number of transpositions producing the identity in S_n. Then we can find such a product that uses the fewest number of transpositions. At least one transposition will involve moving n, since the lemma is true for S_{n-1}. Suppose that the m-th transposition is the first one that moves n. For all possibilities that use the same number of transpositions, we can find one in which m is as large as possible. If only the last transposition moves n, then the product would not be the identity, so there is at least one transposition beyond the m-th. But then the m-th and the $(m + 1)$-th transpositions are one of the four possibilities

$$(n\,x)(n\,x), \qquad (n\,x)(n\,y), \qquad (n\,x)(x\,y), \quad \text{or} \quad (n\,x)(y\,z)$$

for some x, y, and z. In the first case, the two transpositions cancel, so we can form a product using a fewer number of transpositions. In the other three cases, we can replace the pair with another pair,

$$(n\,x)(n\,y) = (x\,y)(n\,x); \quad (n\,x)(x\,y) = (x\,y)(n\,y); \quad (n\,x)(y\,z) = (y\,z)(n\,x);$$

for which m is larger. In all cases, we violate minimality, so there is no odd product of transpositions in S_n equaling the identity. □

We can use this lemma to prove the following theorem.

THEOREM 5.1: The Signature Theorem
For the symmetric group S_n, define the function

$$\sigma : S_n \to \mathbb{Z}$$

by

$$\sigma(x) = (-1)^{N(x)},$$

where $N(x)$ is the minimum number of transpositions needed to express x as a product of transpositions. Then this function, called the signature function*, is a homomorphism from S_n to the set of integers $\{-1, 1\}$.*

PROOF By lemma 5.2, every element of S_n can be written as a product of transpositions, so $\sigma(x)$ is well defined. Obviously this maps S_n to $\{-1, 1\}$, so we only need to establish that this is a homomorphism. Suppose that $\sigma(x \cdot y) \neq \sigma(x) \cdot \sigma(y)$. Then $N(x \cdot y) - (N(x) + N(y))$ would be an odd number. Since $N(x^{-1}) = N(x)$, we would also have $N(x \cdot y) + N(y^{-1}) + N(x^{-1})$ being an odd number. But then we would have three sets of transpositions, totaling an odd number, which when strung together produce $x \cdot y \cdot y^{-1} \cdot x^{-1} = (\,)$. But this contradicts lemma 5.3, so in fact $\sigma(x \cdot y) = \sigma(x) \cdot \sigma(y)$ for all x and y in S_n. □

Because of the importance of the signature function, it is not surprising that *Mathematica* has the signature function built in. For example, the signature of $P[4,3,5,1,2]$ is given by

Signature[P[4,3,5,1,2]]

The signature of an r-cycle will be -1 if r is even, and $+1$ if r is odd. Unfortunately, *Mathematica* will not compute the signature of a cycle directly. One must first convert the cycle or product of cycles into a permutation using the command CycleToPerm, and then compute the signature. Thus, the signature of the product of cycles

$$(1427)(673)$$

is given by

Signature[CycleToPerm[C[1,4,2,7] . C[6,7,3]]]

The corresponding function in GAP is SignPerm:

```
gap> SignPerm( (1,4,2,7)*(6,7,3) );
-1
```

DEFINITION 5.4 A permutation is an *alternating permutation* or an *even permutation* if the signature of the permutation is 1. A permutation is an *odd permutation* if it is not even, that is, if the signature is -1. The set of all alternating permutations of order n is written A_n.

COROLLARY 5.1
The set of all alternating permutations A_n is a normal subgroup of S_n. If $n > 1$, then S_n/A_n is isomorphic to Z_2.

PROOF Clearly A_n is a normal subgroup of S_n, since A_n is the kernel of the signature homomorphism. Also if $n > 1$, then S_n contains at least one transposition whose signature would be -1. Thus, the image of the homomorphism is $\{-1, 1\}$. This group is isomorphic to Z_2. Then by the first isomorphism theorem (4.1), S_n/A_n is isomorphic to Z_2. □

PROPOSITION 5.1
For $n > 2$, the alternating group A_n is generated by the set of 3-cycles.

PROOF Since every 3-cycle is a product of two transpositions, every 3-cycle is in A_n. Thus, it is sufficient to show that every element in A_n can be expressed in terms of 3-cycles. We have already seen that any element can

be expressed as a product of an even number of transpositions. Suppose we group these in pairs as follows:

$$x = [(i_1 j_1) \cdot (k_1 l_1)] \cdot [(i_2 j_2) \cdot (k_2 l_2)] \cdot \cdots \cdot [(i_n j_n) \cdot (k_n l_n)].$$

If we could convert each pair of transpositions into 3-cycles, we would have the permutation x expressed as a product of 3-cycles. There are three cases to consider:

Case 1:

The integers i_m, j_m, k_m, l_m are all distinct. In this case,

$$(i_m j_m) \cdot (k_m l_m) = (i_m j_m l_m) \cdot (i_m k_m l_m).$$

Case 2:

Three of the four integers i_m, j_m, k_m, l_m are distinct. The four combinations that would produce this situation are $i_m = k_m$, $i_m = l_m$, $j_m = k_m$, or $j_m = l_m$. However, these four possibilities are essentially the same, so we only have to check one of these four combinations: $i_m = k_m$. Then we have

$$(i_m j_m) \cdot (i_m l_m) = (i_m j_m l_m).$$

Case 3:

Only two of the four integers i_m, j_m, k_m, and l_m are distinct. Then we must either have $i_m = k_m$ and $j_m = l_m$, or $i_m = l_m$ and $j_m = k_m$. In either case, we have

$$(i_m j_m) \cdot (k_m l_m) = ( ) = (123)(132).$$

In all three cases, we were able to express a pair of transpositions in terms of a product of one or two 3-cycles. Therefore, the permutation x can be written as a product of 3-cycles. □

Let us use this proposition to find the elements of A_4. We know that this is generated by 3-cycles, and has $4!/2 = 12$ elements. Since

Group[{C[1,2,3], C[1,2,4]}]

```
gap> List(Group( (1,2,3),(1,2,4) ) );
[ (), (1,3)(2,4), (1,2)(3,4), (1,4)(2,3), (2,3,4), (1,3,2),
  (1,2,4), (1,4,3), (2,4,3), (1,3,4), (1,2,3), (1,4,2) ]
```

has 12 elements, this must be A_4. Eight of the twelve elements are 3-cycles. The other four elements form a subgroup that we have seen before.

5.3 Cayley's Theorem

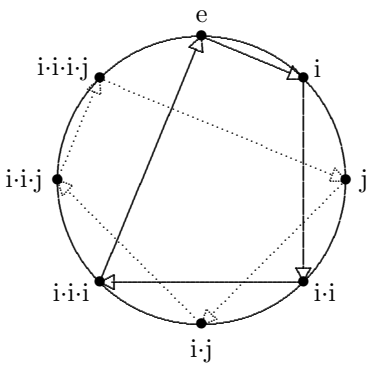
The circle graphs produced in section 5.1 demonstrated the property that every permutation was *one-to-one* and *onto.* In fact, every one-to-one and onto function on a finite set can be seen as a permutation on that set. For example, we saw one-to-one and onto circle graphs in section 3.1 while working with cosets. To demonstrate, let us work with the group Q of order 8:

InitGroup[e];
Define[i^4, e]
Define[j^2, i^2]
Define[j.i, i.i.i.j]
Q = Group[{i, j}]

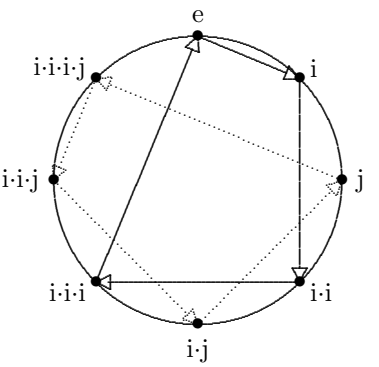
To find the left and right cosets of a subgroup generated by i, we use the commands

CircleGraph[Q, RightMult[i]]
CircleGraph[Q, LeftMult[i]]

which produce the two circle graphs in figure 5.3.



CircleGraph[Q, RightMult[i]]



CircleGraph[Q, LeftMult[i]]

FIGURE 5.3: Circle graphs for multiplying by i

If we number the elements of Q from 1 to 8, starting with e and going clockwise around the circles of figure 5.3, we find that the left circle graph mimics the permutation $P[2,4,5,6,7,1,8,3] = (1246)(3578)$, while the second graph is similar to the permutation $P[2,4,8,6,3,1,5,7] = (1246)(3875)$. If we used different elements of Q in place of the i, we would have a different set of permutations. Thus, we can define two functions, $\phi(x)$ and $f(x)$, which map elements of Q to S_8. Table 5.2 shows both of these two functions.

TABLE 5.2: Permutations for Q

x	$\phi(x)$ RightMult[x]	$f(x)$ LeftMult[x]
e	()	()
i	(1246)(3578)	(1246)(3875)
j	(1347)(2865)	(1347)(2568)
i^2	(14)(26)(37)(58)	(14)(26)(37)(58)
$i \cdot j$	(1548)(2367)	(1548)(2763)
i^3	(1642)(3875)	(1642)(3578)
$i^2 \cdot j$	(1743)(2568)	(1743)(2865)
$i^3 \cdot j$	(1845)(2763)	(1845)(2367)

Let us use GAP to see if either of these is a homomorphism. First we have to define both Q and S_8 in GAP.

```
gap> f := FreeGroup("i","j");; i:=f.1;; j:=f.2;;
gap> Q := f/[i^4,j^2/i^2,j*i/(i^3*j)];; i:=Q.1;; j:=Q.2;;
gap> S8 := SymmetricGroup(8);
Sym( [ 1 .. 8 ] )
gap> Size(S8);
40320
```

Notice that the GAP command `SymmetricGroup` automatically defines S_8. We are now ready for the homomorphism.

```
gap> phi := GroupHomomorphismByImages(Q,S8,[i,j],
> [(1,2,4,6)(3,5,7,8), (1,3,4,7)(2,8,6,5)]);
[ i, j ] -> [ (1,2,4,6)(3,5,7,8), (1,3,4,7)(2,8,6,5) ]
gap> Image(phi, i*j);
(1,8,4,5)(2,7,6,3)
```

So although this produces a homomorphism, it isn't ϕ, since it maps $i \cdot j$ to $(1\,8\,4\,5)(2\,7\,6\,3)$ instead of $(1\,5\,4\,8)(2\,3\,6\,7)$. So ϕ must not be a homomorphism. Let us try seeing if f is a homomorphism.

```
gap> F := GroupHomomorphismByImages(Q,S8,[i,j],
> [(1,2,4,6)(3,8,7,5), (1,3,4,7)(2,5,6,8)]);
[ i, j ] -> [ (1,2,4,6)(3,8,7,5), (1,3,4,7)(2,5,6,8) ]
gap> Image(F,i*j);
(1,5,4,8)(2,7,6,3)
gap> Image(F,i^3*j);
(1,8,4,5)(2,3,6,7)
```

This time, $f(i \cdot j)$ and $f(i^3 \cdot j)$ is exactly the permutation produced by **LeftMult**. So f is a homomorphism, even though ϕ is not. We can easily generalize this to prove the following.

THEOREM 5.2: Cayley's Theorem
Every finite group of order n is isomorphic to a subgroup of S_n.

PROOF Let G be a group of order n. For each g in G, define the mapping

$$p_g : G \to G$$

by $p_g(v) = v \cdot g$. For a given g, if $p_g(v) = p_g(w)$, then $v \cdot g = w \cdot g$, so $v = w$. Hence, p_g is a one-to-one mapping. Also,

$$p_g(v \cdot g^{-1}) = v \cdot g^{-1} \cdot g = v.$$

So every element of G is mapped by an element of G. Thus, p_g is also an onto mapping, and hence is a permutation of the elements of G.

We now can consider the mapping ϕ from G to the symmetric group $S_{|G|}$ on the elements of G, given by

$$\phi(g) = p_g$$

Now, consider two elements $\phi(x)$ and $\phi(y)$. The product of these is the mapping

$$v \to p_y(p_x(v)) = p_y(v \cdot x) = (v \cdot x) \cdot y = v \cdot (x \cdot y).$$

Since this is the same as $\phi(x \cdot y)$, ϕ is a homomorphism.

The element x will be in the kernel of the homomorphism ϕ only if $\phi_x(v)$ is the identity permutation. This means that $v \cdot x = v$ for all elements v in G. Thus, the kernel consists just of the identity element of G, and hence ϕ is an isomorphism. Therefore, G is isomorphic to a subgroup of $S_{|G|}$. □

There is a GAP command `IsomorphismPermGroup` that applies Cayley's theorem to any finite group.

```
gap> iso := IsomorphismPermGroup(Q);
[ i, j ] -> [ (1,2,6,3)(4,8,5,7), (1,4,6,5)(2,7,3,8) ]
gap> Image(iso, i*j);
(1,7,6,8)(2,5,3,4)
```

The slight difference between this isomorphism and the first one that we discovered comes from the fact that GAP ordered the elements of Q differently.

Here is another example: the group D_4, whose multiplication table is given in table 4.2 in chapter 4,

```
gap> f := FreeGroup("a","b");; a:=f.1;; b:= f.2;;
gap> D4 := f/[a^2, b^4, a*b*a*b];; a:= D4.1;; b:= D4.2;;
gap> iso := IsomorphismPermGroup(D4);
[ a, b ] -> [ (2,3), (1,2,4,3)]
gap> List(Image(iso));
[ (), (2,3), (1,3,4,2), (1,3)(2,4), (1,4), (1,4)(2,3),
  (1,2)(3,4), (1,2,4,3) ]
```

Although Cayley's theorem (5.2) shows that D_4 is a subgroup of S_8, GAP actually found a subgroup of S_4 containing an isomorphic copy of D_4. How did GAP do this? Let us consider a *non-normal* subgroup of D_4:

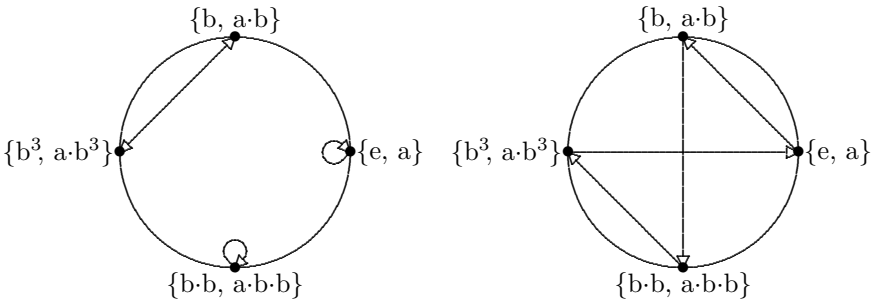
InitGroup[e];
Define[a^2, e]; Define[b^4, e]; Define[b.a, a.b.b.b]
D4 = Group[{a,b}];
H = {e, a}

We saw in Cayley's theorem (5.2) that **LeftMult** applied to the elements of the group derived a homomorphism. What if we applied **LeftMult** to the cosets of the group? Recall that **LeftMult[x]** can be thought of as a function $p_x(v) = v \cdot x$, that is, it multiplies the argument of the function to the left of x. If we apply this function to a right coset of H, we have $p_x(H \cdot g) = H \cdot g \cdot x$, which yields another right coset. (Left cosets won't work here, since $p_x(g \cdot H) = g \cdot H \cdot x$, which is neither a left nor right coset.) The list of right cosets is given by

R = RtCoset[D4, H]

$$\{\{b, a \cdot b\}, \{e, a\}, \{b \cdot b, a \cdot b \cdot b\}, \{b \cdot b \cdot b, a \cdot b \cdot b \cdot b\}\}$$

If we multiply each coset to the left of a fixed element of the group, say a or b, we get the circle graphs in figure 5.4.



CircleGraph[R, LeftMult[a]] **CircleGraph[R, LeftMult[b]]**

FIGURE 5.4: Circle graphs for multiplying cosets of D_4

We see that each coset is mapped to another coset, so once again we can treat each circle graph as a permutation. By numbering the cosets in the order that they appear in **R**, we see that **LeftMult[a]** acts as the permutation $\mathbf{P}[4, 2, 3, 1] = (14)$, whereas **LeftMult[b]** acts as the permutation $\mathbf{P}[3, 1, 4, 2] = (1342)$. *Mathematica* or GAP can check that this extends to a homomorphism.

```
gap> S4 := SymmetricGroup(4);
Sym( [ 1 .. 4 ] )
gap> F := GroupHomomorphismByImages(D4,S4,[a,b],
> [(1,4),(1,3,4,2)]);
[ a, b ] -> [ (1,4), (1,3,4,2) ]
gap> List(Kernel(F));
[ <identity ...> ]
```

Since the kernel is just the identity element, we see that there is a subgroup of S_4 isomorphic to D_4. Note that this is a much stronger result than Cayley's theorem (5.2), which only says that D_4 is isomorphic to a subgroup of the larger group S_8.

PROPOSITION 5.2

Let G be a finite group of order n, and H a subgroup of order m. Then there is a homomorphism from G to S_k, with $k = n/m$, and whose kernel is a subgroup of H.

PROOF Let Q be the set of right cosets $H\backslash G$. For each g in G, define the mapping

$$p_g : Q \to Q$$

by $p_g(H \cdot x) = H \cdot x \cdot g$. Note that this is well defined, since if $H \cdot x = H \cdot y$, then $H \cdot x \cdot g = H \cdot y \cdot g$.

For a given g, if $p_g(H \cdot x) = p_g(H \cdot y)$, then $H \cdot x \cdot g = H \cdot y \cdot g$, so $H \cdot x = H \cdot y$. Hence, p_g is a one-to-one mapping. Also,

$$p_g(H \cdot x \cdot g^{-1}) = H \cdot x \cdot g^{-1} \cdot g = H \cdot x,$$

so every element of Q is mapped by an element of Q. Thus, p_g is also an onto mapping, and hence is a permutation of the elements of Q.

We now can consider the mapping ϕ from G to the symmetric group $S_{|Q|}$ on the elements of Q, given by

$$\phi(g) = p_g.$$

Now, consider two elements $\phi(g)$ and $\phi(h)$. The product of these is the mapping

$$H \cdot x \to p_h(p_g(H \cdot x)) = p_h(H \cdot x \cdot g) = H \cdot x \cdot (g \cdot h).$$

Since this is the same as $\phi(g \cdot h)$, ϕ is a homomorphism.

Finally, we must show that the kernel of ϕ is a subgroup of H. The element g will be in the kernel of the homomorphism ϕ only if $p_g(H \cdot x)$ is the identity permutation. This means that $H \cdot x \cdot g = H \cdot x$ for all right cosets $H \cdot x$ in Q. In particular, the right coset $H \cdot e = H$ is in Q, so $H \cdot g = H$. This can only happen if g is in H. Thus, the kernel is a subgroup of H. We have found a

homomorphism ϕ from the group G to the group $S_{|Q|} = S_k$, whose kernel is a subgroup of H. □

We see one application of this proposition in the case of D_4. Since H was a subgroup of order 2 which was not normal, the only normal subgroup of G that is contained in H is the trivial subgroup. Thus, the homomorphism is an isomorphism, and we find a copy of D_4 inside of S_4 instead of having to look in the larger group S_8. This idea can be applied whenever we can find a subgroup of G that does not contain any nontrivial normal subgroups of G.

But there is another important ramification from this proposition. We can prove the existence of a normal subgroup of a group, knowing only the order of the group!

COROLLARY 5.2

Let G be a finite group, and H a subgroup of G. Then H contains a subgroup N, which is a normal subgroup of G, such that $|G|$ divides $(|G|/|H|)! \cdot |N|$.

PROOF By proposition 5.2, there is a homomorphism ϕ from G to S_k, where $k = |G|/|H|$. Furthermore, the kernel is a subgroup of H. If we let N be the kernel, and let I be the image of the homomorphism, we have by the first isomorphism theorem (4.1) that

$$G/N \approx I.$$

In particular, $|G|/|N| = |I|$, and $|I|$ is a factor of $|S_k| = k!$. This means that $|G|$ is a factor of $k! \cdot |N|$. □

Here is an example of how we can prove the existence of a nontrivial normal subgroup, using just the order of the group. Suppose we have a group G of order 108. Suppose that G has a subgroup of order 27. (We will find in section 7.4 that all groups of order 108 must have a subgroup of order 27.) Using $|G| = 108$ and $|H| = 27$, we find that G must contain a subgroup N such that 108 divides $(108/27)! \cdot |N| = 24 \cdot |N|$. But this means that $|N|$ must be a multiple of 9. Since N is a subgroup of H, which has order 27, we see that N is of order 9 or 27. Hence, we have proven that G contains a normal subgroup of either order 9 or 27. This will go a long way in finding the possible group structures of G, using only the size of the group G.

5.4 Numbering the Permutations

Although using cycles to denote permutations is more succinct in most cases and more readable, *Mathematica* works much faster using the standard permutation notation. Thus, for large time consuming operations, such as checking that a function is a homomorphism, it will be much faster using the $P[\ldots]$ notation than the $C[\ldots]$ notation. For example, we saw using Cayley's theorem that there was a copy of Q inside of S_8. It was generated by the elements

$$f(i) = P[2,4,8,6,3,1,5,7] \qquad \text{and} \qquad f(j) = P[3,5,4,7,6,8,1,2].$$

Thus, we could form a group isomorphic to Q by the command

Q = Group[{P[2,4,8,6,3,1,5,7], P[3,5,4,7,6,8,1,2]}]

Alternatively, we could have used the cycle notation.

Q = Group[{C[1,2,4,6] . C[3,8,7,5] , C[1,3,4,7] . C[2,5,6,8]}]

```
gap> Q := Group( (1,2,4,6)(3,8,7,5), (1,3,4,7)(2,5,6,8) );;
gap> List(Q);
[ (), (1,4)(2,6)(3,7)(5,8), (1,6,4,2)(3,5,7,8),
  (1,2,4,6)(3,8,7,5), (1,7,4,3)(2,8,6,5), (1,3,4,7)(2,5,6,8),
  (1,5,4,8)(2,7,6,3), (1,8,4,5)(2,3,6,7) ]
```

Even though the cycle notation reveals more of the structure of the group (such as the order of each of the elements), it takes *Mathematica* longer to work with cycles. On the other hand, GAP requires working with the cycles notation, since it cannot form a group from transformations.

This section introduces a way to work with permutations in *Mathematica* or GAP that combines succinctness and speed. *Mathematica* has a preset order in which it lists the permutations.

$$\begin{aligned}
\text{1st permutation} &= P[\,] \\
\text{2nd permutation} &= P[2,1] \\
\text{3rd permutation} &= P[1,3,2] \\
\text{4th permutation} &= P[3,1,2] \\
\text{5th permutation} &= P[2,3,1] \\
\text{6th permutation} &= P[3,2,1] \\
\text{7th permutation} &= P[1,2,4,3] \\
&\ldots \quad \ldots \\
\text{24th permutation} &= P[4,3,2,1]
\end{aligned}$$

Notice that the first 2 permutations give the group S_2, the first 6 give S_3, and the first 24 elements give S_4. This pattern can be extended to higher order permutations, so that the first $n!$ permutations gives the group S_n.

The order of the permutations are designed so that *Mathematica* or GAP can quickly find the n-th permutation on the list. For example,

NthPerm[2000]
$P[4, 1, 7, 6, 3, 2, 5]$

```
gap> NthPerm(2000);
(1,4,6,2)(3,7,5)
```

finds the 2000th permutation on this list without having to find the previous 1999. Notice that *Mathematica* returns a permutation, whereas GAP returns the answer in terms of cycles. *Mathematica* and GAP can also quickly determine the position of a given permutation on this list. The command

PermToInt[P[4,1,7,6,3,2,5]]

```
gap> PermToInt( (1,4,6,2)(3,7,5) );
2000
```

converts the permutation back to the number 2000.

Rather than spelling out each permutation, we can now give a single number that describes where the permutation is on the list of permutations. This will be called the *integer representation* of the permutation. Although this representation hides most of the information about the permutation, *Mathematica* and GAP can quickly recover the needed information to do group operations.

For example, we can multiply the 3rd permutation with the 21st on the list with the command

NthPerm[3] . NthPerm[21]

```
gap> NthPerm(3)*NthPerm(21);
(1,2,3,4)
```

If we wanted this converted back to a number, we would type

PermToInt[NthPerm[3] . NthPerm[21]]

```
gap> PermToInt(NthPerm(3)*NthPerm(21));
19
```

Hence the 3rd permutation times the 21st permutation gives the 19th permutation. If we had multiplied in the other order, we would get 23 instead, indicating that the group is non-abelian.

Mathematica provides a shortcut to the previous types of calculations. By entering the command

InitPermMultiplication

we can use the dot product to multiply numbers as if they were permutations. Thus

3 . 21
19

multiplies the 3rd and 21st permutations, and automatically converts this back to a number. Also, the command

23^(-1)
18

finds that the inverse of the 23rd permutation is the 18th permutation. Notice that we need to leave a space between the number and the dot, to distinguish the dot from a decimal point.

This integer representation of the permutations allows us to find other groups within the permutations easily. For example, the quaternion group was generated by the elements

$$P[2,4,8,6,3,1,5,7] \qquad \text{and} \qquad P[3,5,4,7,6,8,1,2].$$

Converting these to integer representations

PermToInt[P[2,4,8,6,3,1,5,7]]
7159
PermToInt[P[3,5,4,7,6,8,1,2]]
34587

we find that the quaternion group can be represented by

TABLE 5.3: Integer representation of Q

$\cdot$	1	7159	12569	18499	23992	25576	34587	37277
1	1	7159	12569	18499	23992	25576	34587	37277
7159	7159	18499	23992	25576	34587	1	37277	12569
12569	12569	37277	18499	34587	7159	23992	1	25576
18499	18499	25576	34587	1	37277	7159	12569	23992
23992	23992	12569	25576	37277	18499	34587	7159	1
25576	25576	1	37277	7159	12569	18499	23992	34587
34587	34587	23992	1	12569	25576	37277	18499	7159
37277	37277	34587	7159	23992	1	12569	25576	18499

InitPermMultiplication
G = Group[{7159, 34587}]
$\{1, 7159, 12569, 18499, 23992, 25576, 34587, 37277\}$

This gives the whole group on a single line which encodes the entire structure of the group. Finally, the command **MultTable[G]** produces table 5.3.

Unfortunately GAP cannot redefine the product of two integers. However, we can still use the succinctness of the integer representation when displaying the multiplication tables by setting the variable `IntPermMultiplication` to `true`.

```
gap> Q := Group( (1,2,4,6)(3,8,7,5), (1,3,4,7)(2,5,6,8) );;
gap> List(Q, x -> PermToInt(x) );
[ 1, 18499, 25576, 7159, 12569, 34587, 37277, 23992 ];
gap> ResetTableOptions();
gap> IntPermMultiplication := true;
true
gap> MultTable(Q);
```

*	1	7159	12569	18499	23992	25576	34587	37277
1	1	7159	12569	18499	23992	25576	34587	37277
7159	7159	18499	23992	25576	34587	1	37277	12569
12569	12569	37277	18499	34587	7159	23992	1	25576
18499	18499	25576	34587	1	37277	7159	12569	23992
23992	23992	12569	25576	37277	18499	34587	7159	1
25576	25576	1	37277	7159	12569	18499	23992	34587
34587	34587	23992	1	12569	25576	37277	18499	7159
37277	37277	34587	7159	23992	1	12569	25576	18499

This integer representation of the permutations allows us to form such a table, and has many other advantages over cyclic permutations, especially when we are working with extremely large subgroups of a symmetric group. Note that the command

```
gap> ResetTableOptions();
```

puts the `MultTable` options back to their default mode.

Problems for Chapter 5

Interactive Problems

5.1 Use *Mathematica* or GAP to find a pair of 3-cycles whose product is a 3-cycle. Can there be a product of two 4-cycles that yields a 4-cycle?

5.2 Use the proof of Cayley's theorem (5.2), with GAP's or *Mathematica*'s help, to find a subgroup of S_8 that is isomorphic to the dihedral group D_4.

5.3 Use Cayley's theorem (5.2) to find a subgroup of S_8 that is isomorphic to Z_{24}^*.

5.4 Find the elements of A_4 converted to the integer representation. Is there a pattern as to which positive integers correspond to the even permutations, and which correspond to odd? Does the pattern continue to A_5?

5.5 Use *Mathematica* or GAP to find all elements of S_7 whose square is $P[3,5,1,7,6,2,4] = (13)(256)(47)$.

Hint: Use a "for" loop to test all of the elements of S_7:

```
For[i = 1, i <= 5040, i++,
  If[ NthPerm[i]^2 == P[3,5,1,7,6,2,4],
    Print[NthPerm[i]]]]
```

In GAP, the corresponding commands are

```
gap> for i in [1..5040] do
>       if ( NthPerm(i)^2 = (1,3)(2,5,6)(4,7) ) then
>          Print( NthPerm(i),"\n" );
>       fi;
>    od;
```

5.6 Use *Mathematica* or GAP to find all elements of S_6 whose cube is $P[3,5,6,1,2,4] = (1364)(25)$. (See the hint for problem 5.5.)

Non-Interactive Problems

5.7 Compute the product

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 1 & 6 & 3 & 2 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 2 & 1 & 6 & 3 & 5 \end{pmatrix}.$$

5.8 Form a multiplication table of S_3 using the permutation notation for the elements. That is, use the elements

$$S_3 = \left\{ \begin{pmatrix} 1\,2\,3 \\ 1\,2\,3 \end{pmatrix}, \begin{pmatrix} 1\,2\,3 \\ 1\,3\,2 \end{pmatrix}, \begin{pmatrix} 1\,2\,3 \\ 2\,1\,3 \end{pmatrix}, \begin{pmatrix} 1\,2\,3 \\ 2\,3\,1 \end{pmatrix}, \begin{pmatrix} 1\,2\,3 \\ 3\,1\,2 \end{pmatrix}, \begin{pmatrix} 1\,2\,3 \\ 3\,2\,1 \end{pmatrix} \right\}.$$

5.9 Find the six elements of S_4 that are of order 4.

Hint: All four of the numbers must move.

5.10 Find a nontrivial element of S_5 that commutes with the permutation

$$x = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 3 & 5 & 1 \end{pmatrix}.$$

5.11 Find a permutation x in S_4 that solves the equation

$$x \cdot \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix} \cdot x.$$

(There are in fact three different answers.)

For problems **5.12** through **5.15**: Find the product of the cycles without using GAP or *Mathematica*.

5.12 $(1\,5\,6) \cdot (3\,5\,2\,4) \cdot (1\,4\,3\,5)$

5.13 $(2\,4\,7) \cdot (1\,3\,6\,4) \cdot (1\,7\,5\,3\,6)$

5.14 $(1\,7\,2\,3\,8\,4) \cdot (1\,3\,5\,2\,4\,6) \cdot (2\,4\,3\,5\,8)$

5.15 $(1\,9\,3\,5\,2\,4\,8) \cdot (2\,7\,3\,9\,5\,4) \cdot (4\,7\,6\,8)$

5.16 Simplify the product of the cycles

$$(1\,3\,2)(2\,4\,3)(3\,5\,4)\cdots(n-1 \;\; n+1 \;\; n)(n \;\; n+2 \;\; n+1)$$

for $n > 1$.

Hint: Try it with $n = 2$, $n = 3$, and $n = 4$ to see a pattern. Then prove using induction that the pattern persists.

5.17 Find the order of the permutations

$$(1\,2\,5)(3\,4) \qquad \text{and} \qquad (1\,2\,5)(3\,4\,6\,7).$$

5.18 Prove that the order of a permutation written in disjoint cycles is the least common multiple of the orders of the cycles.

5.19 Show that A_8 contains an element of order 15.

Hint: See problem 5.18.

5.20 Show that if H is a subgroup of S_n, then either every member of H is an even permutation or exactly half of them are even.

5.21 Find a subgroup of S_4 that is isomorphic to Z_8^*.

Hint: Look at the proof of Cayley's theorem (5.2).

5.22 Find a subgroup of S_5 that is isomorphic to Z_5. (Do you really need Cayley's theorem (5.2) for this one?)

5.23 According to Cayley's theorem (5.2), the quaternion group Q is isomorphic to a subgroup of S_8. Show that Q is not isomorphic to a subgroup of S_7.

Hint: Assume that a subgroup is isomorphic to Q. Is the permutation corresponding to $-1 = i^2$ odd or even? How many disjoint cycles can it contain? What possible permutations can i, j, k, $-i$, $-j$, and $-k$ be mapped to? From this, produce a contradiction.

5.24 In the text we found a group isomorphic to D_4 actually contained in S_4, which is a much smaller group than S_8 used by Cayley's theorem (5.2). What is the smallest symmetric group that contains a subgroup isomorphic to Z_{24}^*?

5.25 *Mathematica* views the permutations

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 3 & 5 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$$

as being the same permutation, $P[2, 1, 4, 3]$. But are these really the same? If not, why can *Mathematica* use the same notation for these two elements?

5.26 The function $\phi(x)$ defined in section 5.3, which used **RightMult** instead of **LeftMult**, was seen not to be a homomorphism. Show that

$$\phi(x \cdot y) = \phi(y) \cdot \phi(x).$$

A function with this property is called an *anti-homomorphism.*

5.27 Let S_Ω be the collection of all one-to-one and onto functions from $\mathbb{Z}^+$ to $\mathbb{Z}^+$ that only move a finite number of elements. Prove that S_Ω is a group. Show that we can write

$$S_\Omega = \bigcup_{n=1}^{\infty} S_n.$$

How should we interpret this union?

5.28 Let S_∞ be the collection of all one-to-one and onto functions from $\mathbb{Z}^+$ to $\mathbb{Z}^+$. Prove that S_∞ is a group. Find an element of this group that is not in S_Ω. (See problem 5.27.)

5.29 Consider the set G of all one-to-one and onto functions $f(x)$ from $\mathbb{Z}^+$ to $\mathbb{Z}^+$ such that there is some integer M for which

$$|f(x) - x| < M \qquad \forall x \in \mathbb{Z}^+.$$

(The value of M is different for different elements of the group.) Prove that G is a group containing S_Ω. Find an element of G that is not in S_Ω. Find an element of S_∞ that is not in G. (See problems 5.27 and 5.28.)

5.30 Show that if G is a group of order 35, and H is a subgroup of order 7, then H is normal.

Hint: Use corollary 5.2.

5.31 Use corollary 5.2 to show that if G is a group of order $p \cdot m$, where p is prime and $p > m$, then any subgroup of order p is normal.

5.32 Let G be a group, and H be a subgroup containing exactly 1/3 of the elements of G. Use corollary 5.2 to show that either H is normal, or exactly half the elements of H form a normal subgroup of G.

5.33 How many elements of order 5 are there in S_6?

5.34 A card-shuffling machine will always shuffle cards in the same way relative to the order in which they were given. All of the spades arranged in order from ace to king are put into the machine, and then the shuffled cards are re-entered into the machine again. If the cards after the second shuffle are in the order 10, 9, 4, Q, 6, J, 5, 3, K, 7, 8, 2, A, what order were the cards in after the first shuffle?

5.35 A subgroup H of the group S_n is called *transitive* on $B = \{1, 2, \ldots, n\}$ if for each pair i, j of elements of B, there exists an element f in H such that $f(i) = j$. Show that there exists a cyclic subgroup H of S_n that is transitive on B.

5.36 Let ϕ denote an r-cycle in S_n, and let x be any permutation in S_n. Show that $x^{-1} \cdot \phi \cdot x$ is an r-cycle.

5.37 Let ϕ and f denote two disjoint cycles in S_n, and let x be any permutation in S_n. Show that $x^{-1} \cdot \phi \cdot x$ and $x^{-1} \cdot f \cdot x$ are disjoint cycles. (See problem 5.36.)

Chapter 6

Building Larger Groups from Smaller Groups

6.1 The Direct Product

In this chapter, we will use the smaller groups that we have previously studied as building blocks to form larger groups. We will discover that *all* finite abelian groups can be constructed using just the cyclic groups Z_n.

One way in which we can create a larger group from two smaller groups is to consider ordered pairs (g_1, g_2), in which the first component g_1 is a member of one group, and the second component g_2 is an element of a second group. We then can multiply these ordered pairs component-wise.

DEFINITION 6.1 Given two groups H and K, the *direct product* of H and K, denoted $H \times K$, is the group of ordered pairs (h, k) such that $h \in H$ and $k \in K$, with multiplication defined by

$$(h_1, k_1) \cdot (h_2, k_2) = (h_1 \cdot h_2, k_1 \cdot k_2).$$

The four group properties for the direct product are easy to verify. Certainly $H \times K$ is closed under multiplication, since the component-wise product of two ordered pairs is again an ordered pair. If e_1 is the identity element for H, and e_2 the identity element for K, then (e_1, e_2) would be the identity element of the direct product. Also, the inverse of an ordered pair (h, k) is (h^{-1}, k^{-1}). Finally, the associative law would hold for $H \times K$, since it holds for both H and K.

Example 6.1

Let $H = Z_4$ and $K = Z_2$. Consider the direct product $G = Z_4 \times Z_2$. Since Z_4 consists of the elements $\{0, 1, 2, 3\}$ and Z_2 consists of $\{0, 1\}$, the set of all ordered pairs (h, k) with $h \in Z_4$ and $k \in Z_2$ is

$$\{(0,0), (0,1), (1,0), (1,1), (2,0), (2,1), (3,0), (3,1)\}.$$ □

The multiplication table for $Z_4 \times Z_2$ is given in *Mathematica*® by the following group of commands:

```
InitGroup[0];
{x_, y_}.{z_, w_} := {Mod[x+z, 4], Mod[y+w, 2]}
MultTable[ {{0,0},{0,1},{1,0},{1,1},{2,0},{2,1}, {3,0},{3,1}}]
```

(Note that curly braces are used here instead of parentheses.) In GAP, we have to first define Z_4 and Z_2 separately, and then form the direct product.

```
gap> f:=FreeGroup("a");; a := f.1;;
gap> Z4 := f/[a^4];;
gap> Z2 := f/[a^2];;
gap> G := DirectProduct(Z4,Z2);
<fp group on the generators [ f1, f2 ]>
gap> List(G);
[ <identity ...>, f1, f2, f1^2, f1*f2, f1^3, f1^2*f2, f1^3*f2 ]
gap> NumberElements := true;;
gap> MultTable(G);
*       |1  2  3  4  5  6  7  8
--------+------------------------
e       |1  2  3  4  5  6  7  8
f1      |2  4  5  6  7  1  8  3
f2      |3  5  1  7  2  8  4  6
f1^2    |4  6  7  1  8  2  3  5
f1*f2   |5  7  2  8  4  3  6  1
f1^3    |6  1  8  2  3  4  5  7
f1^2*f2 |7  8  4  3  6  5  1  2
f1^3*f2 |8  3  6  5  1  7  2  4
```

Notice that GAP picks `f1` and `f2` as the generators of this new group. As a result, the multiplication table is slightly too large to display unless we set the `NumberElements` to true. Nonetheless, we see that this group of eight elements is abelian, has an element of order 4, yet has no element of order 8. Thus by process of elimination, this group must be isomorphic to Z_{15}^*.

PROPOSITION 6.1

Let H and K be two groups. Then $H \times K$ is commutative if, and only if, both H and K are commutative.

PROOF First, suppose that H and K are both commutative. Then for two elements (h_1, k_1) and (h_2, k_2) in $H \times K$, we have

$$(h_1, k_1) \cdot (h_2, k_2) = (h_1 \cdot h_2, k_1 \cdot k_2) = (h_2 \cdot h_1, k_2 \cdot k_1) = (h_2, k_2) \cdot (h_1, k_1).$$

So the two elements in $H \times K$ commute. Hence, $H \times K$ is commutative.

Now suppose that $H \times K$ is commutative. If we let e_1 and e_2 be the identity elements of H and K, respectively, then we have

$$(h_1 \cdot h_2, e_2) = (h_1, e_2) \cdot (h_2, e_2) = (h_2, e_2) \cdot (h_1, e_2) = (h_2 \cdot h_1, e_2)$$

and

$$(e_1, k_1 \cdot k_2) = (e_1, k_1) \cdot (e_1, k_2) = (e_1, k_2) \cdot (e_1, k_1) = (e_1, k_2 \cdot k_1).$$

Thus, $h_1 \cdot h_2 = h_2 \cdot h_1$ and $k_1 \cdot k_2 = k_2 \cdot k_1$ for all h_1 and h_2 in H, and all k_1 and k_2 in K. Hence, both H and K are commutative. □

It is easy to find the number of elements in a direct product. If H has order n, and K has order m, then the number of ordered pairs (h, k) would be $n \cdot m$. We can generalize the direct product to a set of more than two groups. Let

$$G_1, G_2, G_3, \ldots, G_n$$

be a collection of n groups. Then we define $G_1 \times G_2 \times G_3 \times \cdots \times G_n$ to be the set of ordered n-tuples $(g_1, g_2, g_3, \ldots, g_n)$ with multiplication defined by

$$(g_1, g_2, \ldots, g_n) \cdot (h_1, h_2, \ldots, h_n) = (g_1 \cdot h_1, g_2 \cdot h_2, \ldots, g_n \cdot h_n).$$

The direct product of more than two groups can also be defined by taking the direct product of direct products. That is, given three groups G, H, and K, we could define both $(G \times H) \times K$ and $G \times (H \times K)$. But the mappings

$$f : (G \times H) \times K \to G \times H \times K$$

and

$$\phi : G \times (H \times K) \to G \times H \times K$$

given by $f(((g, h), k)) = (g, h, k)$ and $\phi((g, (h, k))) = (g, h, k)$ are clearly surjective isomorphisms. Thus,

$$(G \times H) \times K \approx G \times H \times K \approx G \times (H \times K).$$

It also should be noted that there is the natural mapping

$$f : H \times K \to K \times H$$

given by $f((h, k)) = (k, h)$. This shows that $H \times K \approx K \times H$.

DEFINITION 6.2 Let G be a group. We say that G has a *decomposition* if $G \approx H \times K$, where neither H nor K is the trivial group.

For example, the group Z_{15}^* has a decomposition, since we saw in example 6.1 that this group is isomorphic to $Z_4 \times Z_2$. We would like to find a way of testing whether a general group can be decomposed into smaller groups. The following theorem gives us this test.

THEOREM 6.1: The Direct Product Theorem
Let G be a group with identity e, and let H and K be two subgroups of G. Suppose the following two statements are true:

1. $H \cap K = \{e\}$.

2. *For all* $h \in H$ *and* $k \in K$, $h \cdot k = k \cdot h$.

Then $H \cdot K \approx H \times K$.

PROOF First, let us show that every element in $H \cdot K$ can be uniquely written in the form $h \cdot k$, where $h \in H$ and $k \in K$. Suppose that

$$h_1 \cdot k_1 = h_2 \cdot k_2.$$

Then $h_2^{-1} \cdot h_1 = k_2 \cdot k_1^{-1}$. Since this element must be in both H and K, and the intersection of H and K is the identity element, we have that

$$h_2^{-1} \cdot h_1 = k_2 \cdot k_1^{-1} = e.$$

Thus, $h_1 = h_2$ and $k_1 = k_2$. Therefore, every element of $H \cdot K$ can be written uniquely as $h \cdot k$, where h is in H, and k is in K.

Next, we need to show that $H \cdot K$ is a group. Since $h \cdot k = k \cdot h$ for all $h \in H$ and $k \in K$, we have that $H \cdot K = K \cdot H$. Thus, by lemma 4.4, $H \cdot K$ is a subgroup of G.

We can now define a mapping

$$\phi : H \cdot K \to H \times K$$

by $\phi(x) = (h, k)$, where h and k are the unique elements such that $h \in H$, $k \in K$, and $x = h \cdot k$. It is clear that ϕ is one-to-one, since the element (h, k) can only have come from $h \cdot k$. Also, ϕ is onto, for the element $h \cdot k$ maps to (h, k). All that remains to show that ϕ is an isomorphism is that $\phi(x \cdot y) = \phi(x) \cdot \phi(y)$. Let $x = h_1 \cdot k_1$, and $y = h_2 \cdot k_2$. Then

$$\begin{aligned}
\phi(x \cdot y) &= \phi(h_1 \cdot k_1 \cdot h_2 \cdot k_2) \\
&= \phi(h_1 \cdot h_2 \cdot k_1 \cdot k_2) \\
&= (h_1 \cdot h_2, k_1 \cdot k_2) \\
&= (h_1, k_1) \cdot (h_2, k_2) \\
&= \phi(x) \cdot \phi(y).
\end{aligned}$$

Thus, ϕ is an isomorphism, and so $H \cdot K \approx H \times K$. ∎

We can use this theorem to define the direct product of two groups in *Mathematica*. Suppose we wish to generate the direct product $S_3 \times Z_8^*$. We first must define the two groups in *Mathematica* using the same identity element and different letters for the generators. The group S_3 is defined by the commands

InitGroup[e];
Define[a^3, e]; Define[b^2, e]
Define[b.a, a.a.b]
Define[1/a, a^2]; Define[1/b, b]
H = Group[{a, b}];

Now let us define Z_8^*, using c and d for the two generators.

Define[c^2, e]; Define[d^2, e]
Define[d.c, c.d]
Define[1/c, c]; Define[1/d, d]
K = Group[{c, d}];

Of course we did not use the **InitGroup** command before defining the second group, otherwise we would have cleared the first group. Notice that

Intersection[H, K]

is just the identity element, so the first condition of the direct product theorem is satisfied.

In order for the second condition of the direct product theorem to be satisfied, every element of H must commute with every element of K. This will be true as long as all of the *generators* of H commute with all of the *generators* of K. Since there are 2 generators of H and 2 of K, we can tell *Mathematica* that the generators commute using $2 \cdot 2 = 4$ definitions:

Define[c.a, a.c]; Define[c.b, b.c]
Define[d.a, a.d]; Define[d.b, b.d]

We have to be consistent in the direction of these definitions. That is, we must define an element of the form $k \cdot h$ to $h \cdot k$, where h is in H, and k is in K. This informs *Mathematica* to express all elements as $h \cdot k$.

According to the direct product theorem $H \cdot K$ is now the same as $H \times K$. Here, then, is the direct product:

H . K

Alternatively, we could find the smallest group that contains all of the generators:

G = Group[{a, b, c, d}]

In GAP, we have the option of defining the groups separately, and use the `DirectProduct` that we used for $Z_4 \times Z_2$. But to have more control as to the names of the generators, we can define the direct product as we did in *Mathematica*.

```
gap> f:=FreeGroup("a","b","c","d");;
gap> a := f.1;; b:=f.2;; c:= f.3;; d:=f.4;;
gap> g:= f/[a^3,b^2,b*a*b*a,
> c^2,d^2,c*d*c*d,
> c*a/(a*c),c*b/(b*c),d*a/(a*d),d*b/(b*d)];;
gap> List(g);
[ <identity ...>, b, c, d, a, b*c, b*d, a^2*b, c*d, a*c, a*d,
  a^2, b*c*d, a^2*b*c, a^2*b*d, a*b, a*c*d, a^2*d, a^2*b*c*d,
  a*b*c, a*b*d, a^2*c*d, a*b*c*d ]
gap> Size(g);
24
```

This gives us a group of 24 elements. Since S_4 also has 24 elements, we could ask if these are isomorphic. But recall that S_4 had exactly 9 elements of order 2, whereas the computation

G^2

$\{e,\ a\cdot a,\ e,\ e,\ e,\ a,\ e,\ a\cdot a,\ a\cdot a,\ e,\ e,\ e,\ e,\ a,\ a,\ e,\ e,\ a\cdot a,\ e,\ e,\ e,\ a,\ e,\ e\}$

```
gap> SetReducedMultiplication(g);
gap> List(g, x -> x^2);
[ <identity ...>, <identity ...>, <identity ...>,
  <identity ...>, a^-1, <identity ...>, <identity ...>,
  <identity ...>, <identity ...>, a^-1, a^-1, a, <identity ...>,
  <identity ...>, <identity ...>, <identity ...>, a^-1, a, a,
  <identity ...>, <identity ...>, <identity ...>, a,
  <identity ...> ]
```

reveals that G has 15 elements of order 2. Thus, S_4 is not isomorphic to $S_3 \times Z_8^*$.

This trick of counting the number of solutions to $g^n = e$ for some n is an efficient way of showing that two groups are not isomorphic. We essentially used this with $n = 2$ to show that $S_3 \times Z_8^*$ is not isomorphic to S_4. In fact, it is rather easy to count these solutions for direct products.

PROPOSITION 6.2

Let H and K be finite groups, and let n be a positive integer. If there are r elements of H such that a^n is the identity in H, and s elements of K such that b^n is the identity element of K, then there are $r \cdot s$ elements of $H \times K$ such that c^n is the identity element of $H \times K$.

PROOF Let e_1 denote the identity element of H, and e_2 denote the identity element of K. An element $c = (h, k)$ in $H \times K$ solves the equation $c^n = (e_1, e_2)$ if and only if

$$h^n = e_1 \qquad \text{and} \qquad k^n = e_2.$$

Since there are r solutions to the former, and s solutions to the latter, there are $r \cdot s$ ordered pairs (h, k) that solve both of these equations. Thus, there are $r \cdot s$ elements of $H \times K$ for which $c^n = (e_1, e_2)$. □

For example, there are 4 elements of S_3 satisfying the equation $x^2 = e$, and 4 elements of Z_8^* that satisfy this equation. Thus, there are 16 elements of $S_3 \times Z_8^*$ that satisfy $x^2 = e$, one of which is the identity. Thus, we quickly see that there are 15 elements of order 2.

As powerful as the direct product theorem (6.1) is, it is often difficult to check that $h \cdot k = k \cdot h$ for all $h \in H$ and $k \in K$. Here is a more convenient way of showing that a group can be expressed as a direct product of two subgroups.

COROLLARY 6.1

Let G be a group with identity e, and let H and K be two normal *subgroups of G. Then if $H \cap K = \{e\}$, then $H \cdot K \approx H \times K$.*

PROOF The first condition of the direct product theorem (6.1) is given, so we only need to show that the second condition holds. That is, we need to show that $h \cdot k = k \cdot h$ for all h in H, and k in K. Let $h \in H$ and $k \in K$.

Since K is a normal subgroup of G, $h \cdot k \cdot h^{-1}$ is in K. Thus, $h \cdot k \cdot h^{-1} \cdot k^{-1}$ is in K.

But H is also a normal subgroup of G, so $k \cdot h^{-1} \cdot k^{-1}$ is in H. Hence, $h \cdot k \cdot h^{-1} \cdot k^{-1}$ is in H.

We now use the fact that the only element in both H and K is e. Thus, $h \cdot k \cdot h^{-1} \cdot k^{-1} = e$, which implies $h \cdot k = k \cdot h$. Therefore, the second condition of the direct product theorem (6.1) holds, and so by this theorem, $H \cdot K \approx H \times K$. □

6.2 The Fundamental Theorem of Finite Abelian Groups

In this section, we will show how we can construct any finite commutative group by considering the direct products of the cyclic groups Z_n. We will even be able to find all abelian groups of a given order.

Let us begin with a simple example, Z_6. Can we express this as the direct product of two smaller groups? By the direct product theorem, we must find two subgroups of Z_6 whose intersection is just the identity element, and whose product is the whole group. It is not hard to see that the subgroups

$$H = \{0, 3\} \qquad \text{and} \qquad K = \{0, 2, 4\}$$

satisfy these two conditions. Thus, $Z_6 \approx Z_2 \times Z_3$. This is easily verified using *Mathematica* or GAP. We can first define the group $Z_2 \times Z_3$:

```
gap> f:= FreeGroup("a","b");; a:=f.1;; b:=f.2;;
gap> g:= f/[a^2, b^3, b*a/(a*b)];; a:=g.1;; b:=g.2;;
gap> Order(a*b);
6
gap> StructureDescription(g);
"C6"
```

Since we have an element of order 6, the product $Z_2 \times Z_3$ must be isomorphic to Z_6. GAP's `StructureDescription` command is another way to verify this. GAP uses "C6" instead of Z_6 for the cyclic group of order 6.

Observe the groups $H = \{0, 3\}$ and $K = \{0, 2, 4\}$ in this example. Notice that H consists of all of the elements such that $h^2 = 0$, and K consists of all the elements such that $k^3 = 0$. These two subgroups had only the identity element in common. We can extend this observation to general abelian groups.

LEMMA 6.1
Let G be an abelian group of order mn, where m and n are coprime. Then

$$H = \{h \in G \quad | \quad h^m = e\}$$

and

$$K = \{k \in G \quad | \quad k^n = e\}$$

are both subgroups of G, and $G \approx H \times K$.

PROOF To check that H and K are indeed subgroups simply observe that since G is commutative the functions $\phi(x) = x^m$ and $f(x) = x^n$ are both homomorphisms of G. Then H and K are the kernels of the mappings ϕ and f.

To show that H and K have only the identity element in common, we consider an element x in the intersection. By the Chinese remainder theorem (1.3), there exists a non-negative number $k < m \cdot n$ such that

$$k \equiv 1 \text{ (Mod } m) \qquad \text{and} \qquad k \equiv 0 \text{ (Mod } n).$$

Then $k = (1 + mb)$ for some number b. Thus,

$$x^k = x^{(1+mb)} = x \cdot (x^m)^b = x \cdot e^b = x$$

since x is in H. Yet $k = nc$ for some number c, so

$$x^k = x^{nc} = (x^n)^c = e^c = e$$

since x is in K. Thus, $x = e$, and so $H \cap K = \{e\}$. Since G is abelian, the direct product theorem (6.1) proves that

$$H \cdot K \approx H \times K.$$

All that is left to prove is that $G = H \cdot K$. Let g be an element in G. Since m and n are coprime, by the greatest common divisor theorem (1.2) there exists a and b such that

$$an + bm = \text{GCD}(m, n) = 1.$$

Then

$$g = g^1 = g^{(an+bm)} = g^{an} \cdot g^{bm}.$$

Now, $(g^{an})^m = (g^a)^{nm} = e$, so g^{an} is in H. Likewise, g^{bm} is in K. Thus, every element of G is in $H \cdot K$, and so

$$G \approx H \times K. \qquad \square$$

Unfortunately, the lemma does not tell us that H and K are proper subgroups. It is conceivable that either H or K from lemma 6.1 is the whole group, and the other is just the identity element. We would still have $G = H \times K$, but this would not give a decomposition of G. The next lemma uses induction to show that, in fact, H and K must be nontrivial subgroups.

LEMMA 6.2

If G is a finite abelian group and p is a prime that divides the order of G, then G has an element of order p.

PROOF We will proceed using induction on the order of G. If $|G| = 2$ then p must be 2 and G must isomorphic to Z_2, and so there is an element of order 2 in G.

In fact, whenever $|G|$ is a prime number, then p must be $|G|$, and G must be isomorphic to Z_p. So again, there would be an element of order p in G.

Suppose that the assumption is true for all groups of order less than $|G|$. If G does not have any proper subgroups, then G would be a cyclic group of prime order (which we have already covered.) Thus, we may assume that G has a subgroup N that is neither G nor $\{e\}$.

Since G is abelian all subgroups are normal. Thus we could consider the quotient group G/N. Since $|G| = |N| \cdot |G/N|$, p must divide either $|N|$ or $|G/N|$. If p divides N, then because N is a smaller group than G, by induction N must have an element of order p, which would be in G.

If p does not divide $|N|$ it must divide $|G/N|$. Since G/N is a smaller group than G, by induction G/N must have an element of order p. This element can be written $a \cdot N$ for some a in G.

Since $a \cdot N$ is of order p, a cannot be in N, yet a^p must be in N. If the order of N is q, we would have by corollary 3.2 that $(a^p)^q = e$.

If $b = a^q$ is not the identity, then $b^p = e$, and so b would be the required element. But if $b = e$, then $(a \cdot N)^q = N$. But $a \cdot N$ was of order p, and so

p must divide q. But we assumed that p did not divide $q = |N|$. Hence, b is not the identity, and so G has an element of order p. □

Later on we will see that lemma 6.2 is true for *all* groups, not just abelian groups. However, the result for abelian groups is sufficient for this chapter. This lemma guarantees that the subgroups H and K generated by lemma 6.1 must be proper subgroups. In fact, there are times when it is possible to predict the size of the subgroups H and K.

LEMMA 6.3

Let G be an abelian group of order $p^n \cdot k$ where p is prime, k is not divisible by p, and $n > 0$. Then there are subgroups P and K of G such that $G \approx P \times K$, where $|P| = p^n$, and $|K| = k$.

PROOF Since p^n and k are coprime, we can use lemma 6.1 to form the subgroups

$$P = \{x \in G \mid x^{(p^n)} = e\}$$

and

$$K = \{x \in G \mid x^k = e\}.$$

By lemma 6.1 these two subgroups have only the identity in common, and $G \approx P \times K$. If p divided $|K|$, then by lemma 6.2, K would contain an element of order p. But this element would then be in P as well, which contradicts the fact that only the identity element is in common between P and K. So p does not divide the order of K.

Also note that the order of every element of P is a power of p. Thus, lemma 6.2 tells us that no other prime other than p divides $|P|$.

Finally, note that $|G| = p^n \cdot k = |P| \cdot |K|$. Since p does not divide $|K|$, we have that p^n must divide $|P|$. But no other primes can divide $|P|$, and so $|P| = p^n$. Hence, $|K| = k$. □

Lemma 6.3 is a tremendous help in finding the decomposition of abelian groups. To illustrate, suppose we have an abelian group G of order 24. Since $24 = 2^3 \cdot 3$, lemma 6.3 states that G is isomorphic to a direct product of a group of order 8 and a group of order 3. Thus, G must be one of the groups

$$Z_8 \times Z_3, \qquad Z_{15}^* \times Z_3, \quad \text{or} \quad Z_{24}^* \times Z_3.$$

If we can find all abelian groups of order p^n for p a prime number, then we will in a similar manner be able to find all finite abelian groups.

Hence, our next line of attack is abelian groups of order p^n, where p is prime. If this is not a cyclic group, we can find a decomposition for this group as well.

LEMMA 6.4

Suppose P is an abelian group of order p^n, where p is a prime. Let x be an element in P that has the maximal order of all of the elements of P. Then $P \approx X \times T$, where X is the cyclic group generated by x, and T is a subgroup of P.

PROOF We will use induction on n. If $n = 1$, then P is a cyclic group of order p, and hence is generated by non-identity element x in P. We then have $X = P$, so we can let $T = \{e\}$, and $P \approx X \times T$.

Now suppose that the assertion is true for all powers of p less than n. Notice that the order of every element of P is a power of p. Thus, if we let x be an element with the *largest* order, say m, then the order of all elements in P must divide m. Hence, $g^m = e$ for all elements g in P.

We now let X be the subgroup generated by x. If $X = P$, then we can again let $T = \{e\}$ and we are done. If X is not P, we let y be an element of P not in X which has the *smallest* possible order. Then since the order of y^p is less than the order of y, y^p must be in X. This means that $y^p = x^q$ for some $0 \leq q < m$.

Since y is in P, $y^m = e$. But

$$y^m = (y^p)^{(m/p)} = (x^q)^{(m/p)} = x^{(mq/p)}.$$

Because x is of order m, this can be the identity only if mq/p is a multiple of m. Hence, q is a multiple of p.

If we let $k = x^{-(q/p)} \cdot y$, then k is not in X because y isn't, and

$$k^p = \left(x^{-(q/p)}\right)^p \cdot y^p = x^{-q} \cdot y^p = x^{-q} \cdot x^q = e.$$

Therefore, we have found an element k of order p that is not in X. If we let K be the group generated by the element k, then $X \cap K = \{e\}$.

Consider the quotient group P/K. What is the order of xK in P/K? We see that

$$(xK)^n = K \iff x^n \in K \iff x^n \in X \cap K \iff x^n = e.$$

Therefore, the order of xK is the same as the order of x, which is m. Also note that no element of P/K can have an element of higher order since $a^m = e$ for all elements a in P.

Now we use the induction! Since the order of P/K is less than the order of P, and xK is an element of maximal order, we have by induction that

$$P/K \approx Y \times B,$$

where Y is the subgroup of P/K generated by xK, and B is a subgroup of P/K such that only the identity element K is in the intersection of Y and B.

Let ϕ be the canonical homomorphism from P to P/K given by $\phi(g) = gK$. Let $T = \phi^{-1}(B)$. Then T is a subgroup of P.

If g is in both X and T, then $\phi(g)$ is in both Y and B. Since the intersection of Y and B is the identity element, we have $\phi(g) = g \cdot K = K$. Thus, g is in the subgroup K. But $X \cap K = \{e\}$, so we have

$$X \cap T = \{e\}.$$

Thus, by the direct product theorem (6.1), we find that $X \cdot T \approx X \times T$.

We finally need to show that $P = X \cdot T$. Let u be an element in P, and since $P/K \approx Y \times B$, we can write $\phi(u)$ as $(x^b K) \cdot (kK)$ for some number b, and some kK in B. Then

$$u \in x^b \cdot k \cdot K \subseteq X \cdot T.$$

Thus, $P = X \cdot T$, and so $P \approx X \times T$. □

To illustrate the application of lemma 6.4, consider the group Z_{24}^*. All non-identity elements of Z_{24}^* are of order 2, so this is the maximal order. Thus, lemma 6.4 states that Z_{24}^* can be decomposed into Z_2 and a group of order 4. Since we have seen that $Z_4 \times Z_2 \approx Z_{15}^*$, the only other choice is $Z_2 \times Z_8^*$.

Now we apply lemma 6.4 to Z_8^*. This is of order 4, and all elements besides the identity are of order 2, so Z_8^* can be decomposed into Z_2 and a group of order 2, which must be Z_2. Thus, $Z_8^* \approx Z_2 \times Z_2$, and so

$$Z_{24}^* \approx Z_2 \times Z_2 \times Z_2.$$

We have found a way to decompose any abelian group, as long as its prime decomposition consists of at least two different primes. But now we want to address the issue as to whether a decomposition is *unique.* Can two different decompositions be isomorphic?

The main tool for testing whether two groups are isomorphic is to count elements of a given order. It is natural to ask how many elements there are of a given order for a decomposition of cyclic groups.

LEMMA 6.5

Let p be a prime number, and G be the direct product of cyclic groups

$$Z_{(p^{m_1})} \times Z_{(p^{m_2})} \times \cdots \times Z_{(p^{m_j})} \times Z_{r_1} \times Z_{r_2} \times \cdots \times Z_{r_s},$$

where $m_1, m_2, \ldots, m_j$ are positive integers, and $r_1, r_2, \ldots, r_s$ are coprime to p. Then the number of elements of G of order p^n is given by

$$p^{\left(\sum_{k=1}^{j} \mathrm{Min}(m_k, n)\right)} - p^{\left(\sum_{k=1}^{j} \mathrm{Min}(m_k, n-1)\right)}$$

where $\mathrm{Min}(m_k, n)$ *denotes the minimum of m_k and n.*

PROOF We begin by noticing that the number of elements of order p^n is characterized by the elements y of G such that

$$y^{(p^n)} = e, \quad \text{but} \quad y^{(p^{n-1})} \neq e,$$

where e is the identity of G. Thus, if we can find the number of solutions to the first equation, $y^{(p^n)} = e$, we are on our way to finding the number of elements of order p^n.

Since G is expressed as a direct product we can use proposition 6.2 and find the number of solutions to $y^{(p^n)} = e$ for each factor in the product, and multiply these numbers together. Since each factor is cyclic, we can use corollary 2.1. For all of the factors $Z_{r_1}, Z_{r_2}, \ldots Z_{r_s}$, there is only one solution to $y^{(p^n)} = e$, since $\text{GCD}(r_k, p^n) = 1$. On the other hand, the number of solutions to $y^{(p^n)} = e$ in $Z_{(p^{m_k})}$ is

$$\text{GCD}(p^{m_k}, p^n) = p^{\text{Min}(m_k, n)}.$$

Thus, the number of solutions to $y^{(p^n)} = e$ for the group G is the product of the above for factors 1 through j of G, which gives us a grand total of

$$p^{\left(\sum_{k=1}^{j} \text{Min}(m_k, n)\right)}$$

solutions. However, not all of these solutions will be elements of order p^n. We have to subtract the number of solutions to the equation $y^{(p^{n-1})} = e$ giving us

$$p^{\left(\sum_{k=1}^{j} \text{Min}(m_k, n)\right)} - p^{\left(\sum_{k=1}^{j} \text{Min}(m_k, n-1)\right)}$$

elements of G of order p^n. □

We are now ready to show that *all* finite abelian groups can be represented as the direct product of cyclic groups. However, we would like to show at the same time that such a representation is unique. To this end we will use the previous lemma in conjunction with the following.

LEMMA 6.6

Let $m_1, m_2, m_3, \ldots, m_j$ be a set of positive integers, and define $f(n)$ as

$$f(n) = \sum_{k=1}^{j} \text{Min}(m_k, n)$$

where $\text{Min}(m_k, n)$ denotes the minimum of m_k and n. Then the number of times that the integer n appears in the set of integers $m_1, m_2, m_3, \ldots, m_j$ is given by

$$2f(n) - f(n-1) - f(n+1).$$

PROOF Let us begin by observing the value of the expression

$$2\ \text{Min}(m_k, n) - \text{Min}(m_k, n-1) - \text{Min}(m_k, n+1).$$

When $m_k < n$, then $\text{Min}(m_k, n) = \text{Min}(m_k, n-1) = \text{Min}(m_k, n+1) = m_k$, and so the above evaluates to 0. On the other hand, if $m_k > n$, then the above expression simplifies to be

$$2(n) - (n-1) - (n+1) = 0.$$

However, if $m_k = n$, then $\text{Min}(m_k, n) = n$, $\text{Min}(m_k, n-1) = n-1$, and $\text{Min}(m_k, n+1) = n$. Hence, we have

$$2\ \text{Min}(m_k, n) - \text{Min}(m_k, n-1) - \text{Min}(m_k, n+1) = 2n - (n-1) - n = 1.$$

Thus, we see that

$$2\ \text{Min}(m_k, n) - \text{Min}(m_k, n-1) - \text{Min}(m_k, n+1) = \begin{cases} 1 & \text{if } m_k = n \\ 0 & \text{if } m_k \neq n \end{cases}.$$

Thus, if we sum the above expression for k going from 1 to j, we will count the number of terms m_k that are equal to n. Hence this count will be

$$\sum_{k=1}^{j} 2\ \text{Min}(m_k, n) - \text{Min}(m_k, n-1) - \text{Min}(m_k, n+1) = 2f(n) - f(n-1) - f(n+1).$$

□

We can now use lemmas 6.3 through 6.6 to prove the following.

THEOREM 6.2: The Fundamental Theorem of Finite Abelian Groups

A nontrivial finite abelian group is isomorphic to

$$Z_{(p_1^{m_1})} \times Z_{(p_2^{m_2})} \times Z_{(p_3^{m_3})} \times \cdots Z_{(p_s^{m_s})},$$

where $p_1, p_2, p_3, \ldots, p_s$ are prime numbers (not necessarily distinct). Furthermore, this decomposition is unique up to the rearrangement of the factors.

PROOF We will proceed on induction on the order of the group. If the order of the group is 2, then the theorem is true since the group would be isomorphic to Z_2. Let G be a finite abelian group and suppose the theorem is true for all groups of order less than G. Let p be a prime that divides the order of G. By lemma 6.3, $G \approx P \times K$, where P is the subgroup containing the elements of order p^m for some m.

Furthermore, if x is an element of maximal order in P, and X is the group generated by x, then by lemma 6.4, $G \approx X \times T \times K$. Since X will be a

nontrivial cyclic group the orders of T and K will be less than G. Thus, by induction, T and K can be written as a direct product of cyclic groups whose orders are powers of primes. Since X is also a cyclic group of order p^r, G can be written as a direct product of cyclic groups whose orders are powers of primes.

We next have to show that this decomposition is *unique*. We will do this by showing that the number of times $Z_{(p^n)}$ appears in the decomposition, where p is a prime, is completely determined by the order of the elements in the group G. From lemma 6.5, the number of elements of order p^n is given by

$$p^{(\sum\ \mathrm{Min}(m_k,n))} - p^{(\sum\ \mathrm{Min}(m_k,n-1))}$$

where the sum is taken over all k such that $p_k = p$. Thus, we see that

$$f_p(n) = \sum_{p_k=p} \mathrm{Min}(m_k, n)$$

will be completely determined by the order of the elements of G, and hence determined by the group G. But then by lemma 6.6 the number of times that $Z_{(p^n)}$ appears in the decomposition is given by

$$2f_p(n) - f_p(n-1) - f_p(n+1).$$

Hence, the decomposition of G as a direct product of cyclic groups of the form $Z_{(p^n)}$ is unique. □

From this theorem, we can easily find all non-isomorphic abelian groups of a given order. For example, to find all non-isomorphic abelian groups of order 16, we note that all such groups are direct products of the cyclic groups of orders 2, 4, 8, or 16. This gives us five combinations:

$$Z_2 \times Z_2 \times Z_2 \times Z_2, \qquad Z_2 \times Z_2 \times Z_4, \qquad Z_4 \times Z_4, \qquad Z_2 \times Z_8, \quad \text{and} \quad Z_{16}.$$

Since the fundamental theorem (6.2) also states that the representation is unique, these five groups must be non-isomorphic to each other.

COROLLARY 6.2

Let $P(n)$ denote the number of ways in which n can be expressed as a sum of positive integers, without regard to order. Then if p is a prime number, there are exactly $P(n)$ non-isomorphic abelian groups of order p^n.

PROOF By the fundamental theorem of abelian groups (6.2), every abelian group of order p^n must be isomorphic to

$$Z_{(p^{m_1})} \times Z_{(p^{m_2})} \times Z_{(p^{m_3})} \times \cdots \times Z_{(p^{m_s})}.$$

Also,

$$p^{m_1} \cdot p^{m_2} \cdot p^{m_3} \cdots p^{m_s} = p^n.$$

Hence $m_1 + m_2 + m_3 + \cdots + m_s = n$. Furthermore, the decomposition of the abelian group is unique up to rearrangement of the factors. Thus, there is a one-to-one correspondence between non-isomorphic abelian groups of order p^n and ways n can be written as a sum of positive integers without regard to order. ∎

We call $P(n)$ the number of *partitions* of n. We can have *Mathematica* count the number of partitions for us. For example, to find the number of partitions of the number 4, we can enter

PartitionsP[4]

in *Mathematica*, or

```
gap> NrPartitions(4);
5
```

to find that there are five groups of order 2^4. The number of partitions increases exponentially with n; in fact a *Mathematica* plot reveals that it grows approximately like the function $e^{\sqrt{n}}$.

We can now find the number of non-isomorphic abelian groups of any order.

COROLLARY 6.3
Let $n > 1$ be an integer with prime factorization

$$p_1^{m_1} \cdot p_2^{m_2} \cdot p_3^{m_3} \cdots p_k^{m_k},$$

where $p_1, p_2, p_3, \ldots, p_k$ are distinct primes. Then the number of non-isomorphic abelian groups of order n is given by

$$P(m_1) \cdot P(m_2) \cdot P(m_3) \cdots P(m_k).$$

PROOF We know from the fundamental theorem of abelian groups (6.2) that each such group is isomorphic to a direct product of cyclic groups whose order is a power of a prime. If we collect all factors involving the same primes together, we find that such a group is isomorphic to a direct product of a series of groups of orders $p_1^{m_1}, p_2^{m_2}, p_3^{m_3}, \cdots, p_k^{m_k}$.

We know from corollary 6.2 that there are exactly $P(r)$ non-isomorphic abelian groups of order p^r. Thus, there are $P(m_i)$ possible groups for the i-th factor in this decomposition. Therefore, there are

$$P(m_1) \cdot P(m_2) \cdot P(m_3) \cdots P(m_k)$$

possible ways of forming a product of groups with orders

$$p_1^{m_1}, p_2^{m_2}, p_3^{m_3}, \ldots, p_k^{m_k}.$$

Since the fundamental theorem of abelian groups (6.2) also states that the decomposition is unique up to the rearrangement of the factors, every group thus formed is isomorphically different. So we have exactly $P(m_1) \cdot P(m_2) \cdot P(m_3) \cdots P(m_k)$ non-isomorphic abelian groups of order n. □

For example, suppose we wish to find the number of non-isomorphic abelian groups of order 180 billion. Since $180{,}000{,}000{,}000 = 2^{11} \cdot 3^2 \cdot 5^{10}$, we have that the number of groups is

PartitionsP[11] * PartitionsP[2] * PartitionsP[10]

```
gap> NrPartitions(11) * NrPartitions(2) * NrPartitions(10);
4704
```

giving us 4704 abelian groups of order 180 billion.

6.3 Automorphisms

We have already studied several examples of homomorphisms and isomorphisms *between* two groups, but suppose we considered a mapping from a group *to itself*. For example, we could consider the following mapping from Z_8 onto itself:

DefMultMod[8]
CircleGraph[{0,1,2,3,4,5,6,7}, Mult[3]]

which produces figure 6.1. This mapping could be considered as the permutation

F = P[3, 6, 1, 4, 7, 2, 5]

since the element 0 is left fixed. We can now treat F as a function, and ask whether this is a homomorphism on Z_8. The command

DefSumMod[8]
Z8 = Group[{1}]
CheckHomo[F, Z8]

verifies that F is a homomorphism from Z_8 onto itself.

In GAP, we have to first define a group for which $a^8 = e$. Then we find a map that sends a to a^3.

```
gap> f:=FreeGroup("a");; a := f.1;;
gap> g:=f/[a^8];; a := g.1;;
gap> F := GroupHomomorphismByImages(g,g,[a],[a^3]);
[ a ] -> [ a^3 ]
gap> List(Kernel(F));
[ <identity ...> ]
```

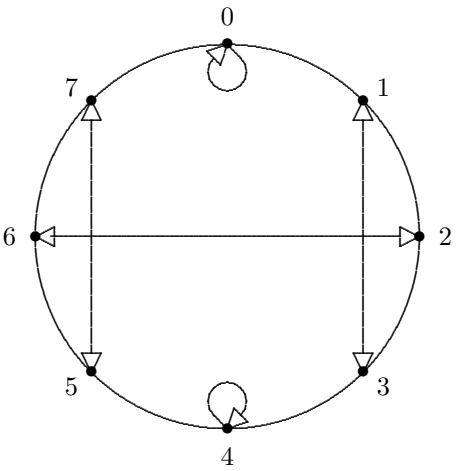


FIGURE 6.1: Multiplying by 3 in Z_8

This shows that in fact the homomorphism is one-to-one and onto.

DEFINITION 6.3 An *automorphism* of the group G is a homomorphism from G to G which is one-to-one and onto.

We can find another automorphism of Z_8 by sending a to a^5 instead of a^3. In fact, it is possible to define the product of two automorphisms as follows: If f and ϕ are both automorphisms of G, then $f \cdot \phi$ is the mapping $x \to \phi(f(x))$. This leads us into the proof of the following.

PROPOSITION 6.3
Given a group G, the set of all automorphisms on G forms a group, denoted $\mathrm{Aut}(G)$. *In fact,* $\mathrm{Aut}(G)$ *is a subgroup of the group of permutations on the elements of G.*

PROOF The mapping $i(x) = x$ for all x in G is obviously an automorphism on G, so the set of all automorphisms on G is non-empty. Also, each automorphism is a permutation on the elements of G. Suppose ϕ and f are two automorphisms on G. Then $\phi(f(x))$ is a one-to-one and onto mapping from G to G.

Furthermore,

$$\phi(f(x \cdot y)) = \phi(f(x) \cdot f(y)) = \phi(f(x)) \cdot \phi(f(y)).$$

So $\phi(f(x))$ is a homomorphism on G, so $f \cdot \phi$ is an automorphism of G.

Also, since f is one-to-one and onto, f^{-1} exists on G, and

$$f\left(f^{-1}(x)\cdot f^{-1}(y)\right) = f\left(f^{-1}(x)\right)\cdot f\left(f^{-1}(y)\right) = x\cdot y.$$

Taking f^{-1} of both sides of the equation gives us

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

So f^{-1} is a homomorphism. So f^{-1}, and thus $\phi\cdot f^{-1}$ are automorphisms of G. Therefore by proposition 2.2, $\mathrm{Aut}(G)$ is a subgroup of the group of permutations on the elements of G. □

Let's see if we can find the automorphism group for Z_8. The element 1 must be mapped by an automorphism to an element of order 8. Thus, 1 is mapped to either 1, 3, 5, or 7. But since 1 is a generator of Z_8, this would completely define the automorphism. Thus, there at most four elements of $\mathrm{Aut}(Z_8)$. But we have already seen three nontrivial automorphisms (multiplying by 3, by 5, and the product of these two), so we have exactly four automorphisms of Z_8. This automorphism group can quickly be seen to be isomorphic to Z_8^*.

GAP can find the automorphism group of Z_8 as follows:

```
gap> f:= FreeGroup("a");; a:=f.1;;
gap> Z8 := f/[a^8];; a:=Z8.1;;
gap> SetReducedMultiplication(Z8);
gap> G := AutomorphismGroup(Z8);
<group with 2 generators>
gap> L := List(G);
[ IdentityMapping( <fp group of size 8 on the generators [a]> ),
  [ a ] -> [ a^3 ], [ a ] -> [ a^-3 ], [ a ] -> [ a^-1 ] ]
```

We see that the automorphism that we first defined is the second one in this list. We can evaluate one of the automorphisms at an element of Z_8 by using the notation `x^f`, where f is the automorphism, as we did for cycles.

```
gap> a^L[2];
a^3
gap> (a^3)^L[3];
a^-1
L[2] * L[3] = L[4];
true
```

This last command shows that we can multiply automorphisms in GAP, and produce another automorphism. This means that we could display a multiplication table of `G`, but each element has a very long name in GAP. A better alternative is to find a permutation group isomorphic to G using the `NiceObject` command.

```
gap> H := NiceObject(G);
Group([ (1,4)(2,3), (1,3)(2,4) ])
gap> ResetTableOptions();
```

```
gap> MultTable(H);
*          |()          (1,4)(2,3)  (1,3)(2,4)  (1,2)(3,4)
----------+----------------------------------------------
()         |()          (1,4)(2,3)  (1,3)(2,4)  (1,2)(3,4)
(1,4)(2,3)|(1,4)(2,3)  ()          (1,2)(3,4)  (1,3)(2,4)
(1,3)(2,4)|(1,3)(2,4)  (1,2)(3,4)  ()          (1,4)(2,3)
(1,2)(2,4)|(1,2)(3,4)  (1,3)(2,4)  (1,4)(2,3)  ()
```

This multiplication table clearly shows that $\mathrm{Aut}(Z_8) \approx Z_8^*$. It is not hard to generalize this result.

PROPOSITION 6.4

$$\mathrm{Aut}(Z_n) \approx Z_n^*.$$

PROOF Consider the mapping

$$\phi : Z_n^* \to \mathrm{Aut}(Z_n)$$

given by $\phi(j) = f_j$, where $f_j(x) = j \cdot x \pmod{n}$. Then given two elements j_1 and j_2 in Z_n^*, we have that

$$f_{j_1}(f_{j_2}(x)) = j_1 \cdot (j_2 x) \pmod{n} = (j_2 \cdot j_1)x \pmod{n} = f_{j_2 \cdot j_1}(x).$$

So

$$\phi(j_2) \cdot \phi(j_1) = f_{j_1} \cdot f_{j_2} = f_{j_1 \cdot j_2} = \phi(j_2 \cdot j_1).$$

Hence, ϕ is a homomorphism from Z_n^* to $\mathrm{Aut}(Z_n)$. To see that ϕ is one-to-one, note that $f_j(1) = j$, and so $f_{j_1} = f_{j_2}$ only if $j_1 = j_2$. To see that ϕ is onto, we consider a general automorphism f of Z_n. Since 1 is a generator of Z_n, $f(1)$ must also be a generator of Z_n. But f will be completely determined by knowing $f(1)$. Thus, the number of automorphisms is at most the number of generators in Z_n^*. But we have an automorphism for each such generator, accounting for all automorphisms of Z_n. □

So far, the automorphism group is smaller than the original group. But let us look at a non-cyclic group, Z_8^*.

InitGroup[e];
Define[a^2,e]; Define[b^2,e]
Define[b.a, a.b]
Define[1/a, a]; Define[1/b, b]
G = Group[{a, b}]

There are in fact six automorphisms of this group. The automorphism

$$f(e) = e$$
$$f(a) = b$$
$$f(b) = a$$
$$f(a \cdot b) = a \cdot b$$

can be represented as a transposition $(a\ b)$. Note that here, we are using the cycle notation with *elements* in place of numbers. This is allowed in *Mathematica*, but not in GAP. So this function can be entered into *Mathematica* simply as

F = C[a, b]

Mathematica can check if this is an automorphism of Z_8^*.

CheckHomo[F, G]

The other automorphisms of Z_8^* can be found quicker in GAP.

```
gap> f:= FreeGroup("a","b");; a:=f.1;; b:=f.2;;
gap> g:= f/[a^2,b^2,b*a/(a*b)];; a:=g.1;; b:= g.2;;
gap> SetReducedMultiplication{g};
gap> G := AutomorphismGroup(g);
<group with 4 generators>
gap> L := List(G);
[ IdentityMapping(<fp group of size 4 on the generators [a,b]>),
  [ a, b ] -> [ a, a^-1*b^-1 ], [ a, b ] -> [ b, a ],
  [ a, b ] -> [ b, a^-1*b^-1 ], [ a, b ] -> [ a^-1*b^-1, a ],
  [ a, b ] -> [ a^-1*b^-1, b ] ]
gap> List(NiceObject(G));
[ (), (1,2,3), (1,3,2), (2,3), (1,2), (1,3) ]
```

The automorphism we found earlier is the third one in this list, and when GAP converts this to a subgroup of a permutation group, we get the six elements of S_3. Hence $\mathrm{Aut}(Z_8^*) \approx S_3$.

For the next example, let us look at the automorphisms for the quaternion group Q.

InitGroup[e];
Define[i^4, e]; Define[j^2, i^2]
Define[j.i, i.i.i.j]
Define[1/i, i^3]; Define[1/j, i.i.j]
Q = Group[{i, j}]

If f is an automorphism of Q, then $f(e) = e$, but also $f(i^2)$ must be i^2, since this is the only element of order 2. All of the other elements are of order 4, so $f(i)$ could be any one of the remaining six elements. Once $f(i)$ is determined, we have that $f(i^3) = f(i)^3$. Then $f(j)$ would be one of the remaining four elements. Since i and j generate Q, f will be determined by knowing $f(i)$ and $f(j)$. Thus, there is a maximum of $6 \cdot 4 = 24$ automorphisms.

For non-commutative groups, there is a quick way to find many of the automorphisms. Let G be a non-commutative group, and let x be any element in G. The mapping $f_x : G \to G$ defined by

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

will always be an automorphism, for

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

Also, f_x is one-to-one and onto, for its inverse is $f_{x^{-1}}$.

DEFINITION 6.4 An automorphism $\phi(y)$ of a group G is called an *inner automorphism* if there is an element x in G such that

$$\phi(y) = x^{-1} \cdot y \cdot x \qquad \text{for all } y \in G.$$

The set of inner automorphisms of G is denoted $\text{Inn}(G)$.

It is fairly easy to find the inner automorphisms on Q. If we choose $x = i$, we have the mapping

$$\begin{array}{ll} f(e) = i^3 \cdot e \cdot i = e & f(i^3) = i^3 \cdot i = i^3 \\ f(i) = i^3 \cdot i \cdot i = i & f(i \cdot j) = i^3 \cdot (i \cdot j) \cdot i = i^3 \cdot j \\ f(j) = i^3 \cdot j \cdot i = i^2 \cdot j & f(i^2 \cdot j) = i^3 \cdot (i^2 \cdot j) \cdot i = j \\ f(i^2) = i^3 \cdot i^2 \cdot i = i^2 & f(i^3 \cdot j) = i^3 \cdot (i^3 \cdot j) \cdot i = i \cdot j \end{array}$$

In GAP, the command `InnerAutomorphism` allows us to enter this mapping.

```
gap> f := FreeGroup("i","j");; i := f.1;; j := f.2;;
gap> Q := f/[i^4, i^2*j^2, j*i/(i^3*j)];; i := Q.1;; j := Q.2;;
gap> SetReducedMultiplication(Q);
gap> F := InnerAutomorphism(Q,i);
^i
gap> j^F;
j^-1
```

In GAP, this inner automorphism is simply referred to as `^i`. This is mainly because GAP uses an abbreviation x^y for $y^{-1} \cdot x \cdot y$.

In *Mathematica*, the automorphism has to be entered as cycles containing the elements of Q.

F = C[j, i.i.j] . C[i.j, i.i.i.j]

If we use $x = j$ or $x = i \cdot j$ instead of $x = i$, we get the automorphisms

G = C[i, i.i.i] . C[i.j, i.i.i.j]
H = C[i, i.i.i] . C[j, i.i.i]

In fact, these three automorphisms, along with the identity automorphism, form a group. These are the only four inner automorphisms.

However, there are many more automorphisms of Q. The commands

Homomorph[X]
Define[X[i], i]
Define[X[j], i.j]
CheckHomo[X, Q]

show that there is another homomorphism from Q to itself, which can be shown to be one-to-one and onto. Also, the commands

Homomorph[J]
Define[J[i], i.j]
Define[J[j], j]
CheckHomo[J, Q]

show that there is yet another automorphism on Q. These two automorphisms, along with the group of 4 previously found, generate a total of 24 automorphisms. We can get all of the automorphisms in GAP as follows:

```
gap> A := AutomorphismGroup(Q);
<group of size 24 with 4 generators>
gap> L := List(A);;
gap> L[2];
[ i^-1, i^-1*j^-1 ] -> [ i^-1, j^-1 ]
gap> L[3];
^i
```

Although there are too many automorphisms to list here, we can notice that the inner automorphisms are embedded in this list. What is this group isomorphic to? We can have GAP provide the answer.

```
gap> StructureDescription(A);
"S4"
```

In fact, $\mathrm{Aut}(Q) \approx S_4$, as can be seen by figure 6.2. Each rotation of the octahedron represents an automorphism of Q. For example, rotating the front face 120° clockwise corresponds to the automorphism

$$(i\ \ j\ \ ij)(i^3\ \ i^2j\ \ i^3j).$$

So the automorphism group is isomorphic to the octahedral group, which we saw was isomorphic to S_4.

Although the inner automorphisms did not produce the full automorphism group, this set of inner automorphisms turns out to be a very important subgroup of the automorphism group. Let us discover the first main property of this subgroup.

PROPOSITION 6.5
Let G be a group. Then $\mathrm{Inn}(G)$ *is a normal subgroup of* $\mathrm{Aut}(G)$.

PROOF First we need to show that $\mathrm{Inn}(G)$ is a subgroup. Let $f_x(y) = x^{-1} \cdot y \cdot x$ be an inner automorphism. The inverse can be easily found by observing

$$y \in f_x^{-1}(v) \Longleftrightarrow x^{-1} \cdot y \cdot x = v \Longleftrightarrow y = x \cdot v \cdot x^{-1} \Longleftrightarrow y = f_{(x^{-1})}(v),$$

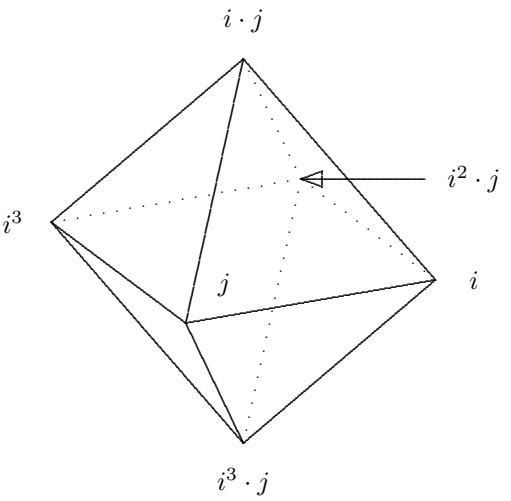


FIGURE 6.2: Labeling the octahedron to show $\mathrm{Aut}(Q)$

so the inverse of f_x is also an inner automorphism.

If we consider two inner automorphisms f_x and f_y, then

$$(f_x \cdot f_y)(v) = f_y(f_x(v)) = y^{-1} \cdot (x^{-1} \cdot v \cdot x) \cdot y = (x \cdot y)^{-1} \cdot v \cdot (x \cdot y) = f_{(x \cdot y)}(v).$$

Thus the product of two inner automorphisms is also an inner automorphism. So by proposition 2.2, $\mathrm{Inn}(G)$ is a subgroup of $\mathrm{Aut}(G)$.

Finally, we need to show that $\mathrm{Inn}(G)$ is normal in $\mathrm{Aut}(G)$. Let ϕ be any automorphism and let $f_x = x^{-1} \cdot y \cdot x$ be an inner automorphism. Then

$$(\phi \cdot f_x \cdot \phi^{-1})(v) = \phi^{-1}(f_x(\phi(v))) = \phi^{-1}(x^{-1} \cdot (\phi(v)) \cdot x).$$

Since ϕ^{-1} is a homomorphism, this will simplify.

$$\begin{aligned} \phi^{-1}(x^{-1} \cdot (\phi(v)) \cdot x) &= \phi^{-1}(x^{-1}) \cdot \phi^{-1}(\phi(v))\phi^{-1}(x) \\ &= (\phi^{-1}(x))^{-1} \cdot v \cdot \phi(x)^{-1} = f_{\phi^{-1}(x)}(v). \end{aligned}$$

So $\phi \cdot f_x \cdot \phi^{-1}$ is an inner automorphism of G. Therefore, by proposition 3.4, $\mathrm{Inn}(G)$ is a normal subgroup of $\mathrm{Aut}(G)$. □

For example, we found four inner-automorphisms of Q. By looking at the multiplication table for these four elements, we see that $\mathrm{Inn}(Q) \approx Z_8^*$.

DEFINITION 6.5 We define the *outer automorphism group* to be the quotient group

$$\mathrm{Out}(G) = \mathrm{Aut}(G)/\mathrm{Inn}(G).$$

The outer automorphism group of Q must contain six elements, and with some experimenting in *Mathematica*, one finds that $\mathrm{Out}(Q)$ is non-abelian. Therefore, $\mathrm{Out}(Q) \approx S_3$.

Let us look at one last example—Z_{24}^*. We can load this group into *Mathematica* with the following commands:

InitGroup[e];
Define[a^2, e]; Define[b^2, e]; Define[c^2, e]
Define[b.a, a.b]; Define[c.a, a.c]; Define[c.b, b.c]
Define[1/a, a]; Define[1/b, b]; Define[1/c, c]
Y = Group[{a, b, c}]

Suppose $\phi(x)$ is an automorphism of Z_{24}^*. Naturally $\phi(e) = e$, but $\phi(a)$ could be any of the seven remaining elements of order 2. Also, $\phi(b)$ could be any one of the remaining six elements. Then we would have $\phi(a \cdot b) = \phi(a) \cdot \phi(b)$. But $\phi(c)$ could be any of the four elements left over. Since the group is generated by $\{a, b, c\}$, there are at most $7 \cdot 6 \cdot 4 = 168$ possible automorphisms.

One possible automorphism would be to send a to b, b to c, and c back to a. This is represented by the permutation

F = C[a, b, c] . C[a.b, b.c, a.c]
CheckHomo[F, Y]

which *Mathematica* verifies is an automorphism. Another automorphism, given by

G = C[b, a.b] . C[b.c, a.b.c]
CheckHomo[G, Y]

indicates that there may indeed be many automorphisms. These can be checked by GAP as follows:

```
gap> f:= FreeGroup("a","b","c");; a:=f.1;; b:=f.2;; c:=f.3;;
gap> g:= f/[a^2,b^2,c^2,a*b*a*b,a*c*a*c,b*c*b*c];;
gap> a:= g.1;; b:=g.2;; c:=g.3;;
gap> F := GroupHomomorphismByImages(g,g,[a,b,c],[b,c,a]);
[ a, b, c ] -> [ b, c, a ]
gap> List(Kernel(F));
[ <identity ...> ]
gap> G := GroupHomomorphismByImages(g,g,[a,b,c],[a, a*b, c]);
gap> List(Kernel(G));
[ <identity ...> ]
gap> A := AutomorphismGroup(g);
<group with 4 generators>
gap> Size(A);
168
```

GAP has indicated that the automorphism group is indeed as large as we had predicted it could be. It would be more concise if we could use permutations for a group this large. If we order the non-identity elements $a = 1$, $b = 2$,

$c = 3$, $a \cdot b = 4$, $a \cdot c = 5$, $b \cdot c = 6$, and $a \cdot b \cdot c = 7$, we can convert F and G to standard permutations $(1,2,3)(4,5,6)$ and $(2,4)(6,7)$. Once we have all of the elements as permutations, we can use the integer notation to list them.

```
gap> A := Group( (1,2,3)(4,6,5), (2,4)(6,7) );
Group([ (1,2,3)(4,6,5), (2,4)(6,7) ])
gap> List(A, x->PermToInt(x) );
[ 1, 244, 149, 918, 2380, 1732, 2002, 735, 2183, 1475, 1649,
  1079, 2471, 3936, 3195, 3817, 4753, 5023, 3595, 4190, 2632,
  1881, 1311, 2847, 4309, 4904, 3476, 3358, 2677, 2107, 1123,
  404, 496, 670, 1432, 3991, 4616, 3032, 2918, 3622, 4384, 4558,
  775, 2240, 1537, 1662, 1014, 2476, 3898, 61, 331, 231, 953,
  2345, 1775, 1992, 1851, 1229, 2787, 4205, 4817, 3372, 3276,
  3177, 3755, 4713, 4931, 3486, 4098, 2562, 3973, 4581, 3019,
  2900, 3662, 4366, 4476, 2647, 2042, 1088, 374, 548, 640, 1362,
  1807, 1202, 2761, 4226, 4874, 3412, 3298, 1837, 1267, 2821,
  4269, 4847, 3455, 3336, 4035, 4657, 3099, 2981, 3689, 4428,
  4498, 4017, 4595, 3059, 2963, 3702, 4410, 4536, 753, 2201,
  1461, 1582, 970, 2418, 3876, 793, 2258, 1496, 1622, 1052,
  2510, 3958, 3151, 3776, 4735, 4970, 3508, 4156, 2602, 3133,
  3741, 4695, 4965, 3573, 4151, 2592, 2691, 2069, 1133, 437,
  593, 684, 1402, 2721, 2151, 1185, 467, 558, 714, 1392, 87,
  357, 187, 908, 2366, 1796, 2032, 27, 270, 122, 856, 2304,
  1692, 1962 ]
```

In *Mathematica*, we can merely note that F is the 149th permutation, and G is the 735th. Thus, we get the same result with the commands

```
InitPermMultiplication
A = Group[{149, 735}]
{1, 27, 61, 87, 122, 149, 187, 231, 244, 270, 331, 357, 374, 404, 437, 467,
 496, 548, 558, 593, 640, 670, 684, 714, 735, 753, 775, 793, 856, 908, 918,
 953, 970, 1014, 1052, 1079, 1088, 1123, 1133, 1185, 1202, 1229, 1267, 1311,
 1362, 1392, 1402, 1432, 1461, 1475, 1496, 1537, 1582, 1622, 1649, 1662, 1692,
 1732, 1775, 1796, 1807, 1837, 1851, 1881, 1962, 1992, 2002, 2032, 2042, 2069,
 2107, 2151, 2183, 2201, 2240, 2258, 2304, 2345, 2366, 2380, 2418, 2471, 2476,
 2510, 2562, 2592, 2602, 2632, 2647, 2677, 2691, 2721, 2761, 2787, 2821, 2847,
 2900, 2918, 2963, 2981, 3019, 3032, 3059, 3099, 3133, 3151, 3177, 3195, 3276,
 3298, 3336, 3358, 3372, 3412, 3455, 3476, 3486, 3508, 3573, 3595, 3622, 3662,
 3689, 3702, 3741, 3755, 3776, 3817, 3876, 3898, 3936, 3958, 3973, 3991, 4017,
 4035, 4098, 4151, 4156, 4190, 4205, 4226, 4269, 4309, 4366, 4384, 4410, 4428,
 4476, 4498, 4536, 4558, 4581, 4595, 4616, 4657, 4695, 4713, 4735, 4753, 4817,
 4847, 4874, 4904, 4931, 4965, 4970, 5023}
```

Notice that *Mathematica* orders the numbers, making it easier to find a particular element. The group $\mathrm{Aut}(Z_{24}^*)$ has some special properties that we will explore in the next chapter.

We have now seen several examples where the group of automorphisms is larger than the original group. But this group of automorphisms can also be used as a tool for connecting two groups to form an even larger group, in much the same way that two groups formed the direct product. The next section will explore this methodology.

6.4 Semi-Direct Products

We have already seen one way to combine two groups H and K to form the direct product $H \times K$. In this section we will see another way to combine to groups H and K. Once again the larger group will have isomorphic copies of H and K as subgroups, but only *one* of the two subgroups will be a normal subgroup.

Suppose that H and K are any two groups, and suppose that we have a homomorphism $\phi : H \to \text{Aut}(K)$. Because the function ϕ returns another function, we will write ϕ_h instead of $\phi(h)$. The expression $\phi_h(k)$ represents the automorphism ϕ_h evaluated at the element k. That is, if h_1 and h_2 are two elements of H, then $\phi_{h_1}(k)$ and $\phi_{h_2}(k)$ will be two automorphisms of K, and also $\phi_{h_1 \cdot h_2}(k) = (\phi_{h_1} \cdot \phi_{h_2})(k) = \phi_{h_2}(\phi_{h_1}(k))$. (Recall that $\phi_{h_1} \cdot \phi_{h_2}$ means we do ϕ_{h_1} first, then do ϕ_{h_2}.)

There will always be at least one homomorphism from H to $\text{Aut}(K)$, the trivial homomorphism. However, there will often be several nontrivial homomorphisms from H to $\text{Aut}(K)$. For each such homomorphism, we can define a product of H and K.

DEFINITION 6.6 Let G be the set of all ordered pairs (h, k), where h is in H and k is in K. Let ϕ be a nontrivial homomorphism from H to $\text{Aut}(K)$. Then the *semi-direct product of K with H through ϕ,* denoted $H \ltimes_{\phi} K$, is the set G with multiplication defined by

$$(h_1, k_1) \cdot (h_2, k_2) = (h_1 \cdot h_2,\ \ \phi_{h_2}(k_1) \cdot k_2).$$

PROPOSITION 6.6
The semi-direct product of K with H through ϕ is a group.

PROOF It is clear that the product of two ordered pairs in G is an ordered pair in G. If we let e_1 denote the identity element of H, and e_2 denote the identity element of K, then

$$\phi_{e_1}(k_1) = k_1,$$

since ϕ must map e_1 to the identity automorphism of K. Thus

$$(h_1, k_1) \cdot (e_1, e_2) = (h_1 \cdot e_1, \phi_{e_1}(k_1) \cdot e_2) = (h_1, k_1),$$

and

$$(e_1, e_2) \cdot (h_2, k_2) = (e_1 \cdot h_2, \phi_{h_2}(e_2) \cdot k_2) = (h_2, k_2).$$

So (e_1, e_2) acts as the identity element of G.

Next we note that the element (h, k) has an inverse $(h^{-1}, \phi_{h^{-1}}(k^{-1}))$, since

$$\begin{aligned}(h^{-1}, \phi_{h^{-1}}(k^{-1})) \cdot (h, k) &= (h^{-1} \cdot h, \phi_h(\phi_{h^{-1}}(k^{-1})) \cdot k) \\ &= (e_1, \phi_{e_1}(k^{-1}) \cdot k) = (e_1, k^{-1} \cdot k) = (e_1, e_2),\end{aligned}$$

and

$$\begin{aligned}(h, k) \cdot (h^{-1}, \phi_{h^{-1}}(k^{-1})) &= (h \cdot h^{-1}, \phi_{h^{-1}}(k) \cdot \phi_{h^{-1}}(k^{-1})) \\ &= (e_1, \phi_{h^{-1}}(k \cdot k^{-1})) = (e_1, \phi_{h^{-1}}(e_2)) = (e_1, e_2).\end{aligned}$$

The final thing we need to check is that the multiplication on G is associative. Note that

$$\begin{aligned}(h_1, k_1) \cdot [(h_2, k_2) \cdot (h_3, k_3)] &= (h_1, k_1) \cdot (h_2 \cdot h_3, \phi_{h_3}(k_2) \cdot k_3) \\ &= (h_1 \cdot h_2 \cdot h_3, \phi_{h_2 \cdot h_3}(k_1) \cdot \phi_{h_3}(k_2) \cdot k_3)\end{aligned}$$

while

$$\begin{aligned}[(h_1, k_1) \cdot (h_2, k_2)] \cdot (h_3, k_3) &= (h_1 \cdot h_2, \phi_{h_2}(k_1) \cdot k_2) \cdot (h_3, k_3) \\ &= (h_1 \cdot h_2 \cdot h_3, \phi_{h_3}(\phi_{h_2}(k_1) \cdot k_2) \cdot k_3) \\ &= (h_1 \cdot h_2 \cdot h_3, \phi_{h_3}(\phi_{h_2}(k_1)) \cdot \phi_{h_3}(k_2) \cdot k_3) \\ &= (h_1 \cdot h_2 \cdot h_3, \phi_{h_2 \cdot h_3}(k_1) \cdot \phi_{h_3}(k_2) \cdot k_3)\end{aligned}$$

Hence the multiplication on G is associative and so G forms a group. □

We can define a semi-direct group in GAP using the definition. Suppose that we wish to find a semi-direct product of the form $Z_4 \ltimes_\phi Z_8^*$. The first step is to define both Z_4 and Z_8^*. We will use a for the generator of Z_4, and b and c for the generators of Z_8^*.

```
gap> f:= FreeGroup("a");; a:=f.1;;
gap> Z4 := f/[a^4];; a := Z4.1;;
gap> f:=FreeGroup("b","c");; b:=f.1;; c:=f.2;;
gap> g:=f/[b^2,c^2,b*c*b*c];;
```

Now we find the automorphism group of Z_8^*.

```
gap> A := AutomorphismGroup(g);
<group with 4 generators>
gap> L := List(A);
[ IdentityMapping(<fp group of size 4 on the generators [b,c]>),
  [ b^-1, c^-1 ]->[ b^-1, c*b ], [ b^-1, c^-1 ]->[ c^-1, b^-1 ],
  [ b^-1, c^-1 ]->[ c^-1, b^-1*c^-1 ],
  [ b^-1, c^-1 ]->[ c*b, b^-1 ],
  [ b^-1, c^-1 ]->[b^-1*c^-1, c^-1] ]
```

A homomorphism that maps Z_4 to this group must send the identity element to an element of order 2 or 4, but $\mathrm{Aut}(Z_8^*)$ has only six elements, so we must find one of order 2. The third element in this list will do, since it exchanges b^{-1} and c^{-1}.

```
gap> phi := GroupHomomorphismByImages(Z4,A,[a],[L[3]]);
[ a ] -> [ [ b^-1, c^-1 ] -> [ c^-1, b^-1 ] ]
```

Notice that we now have a mapping that sends elements of Z_4 to *mappings*. With this, we can define the semi-direct product $Z_4 \ltimes_\phi Z_8^*$ with the commands

```
gap> Size(Z4);
4
gap> S:= SemidirectProduct(Z4,phi,g);
<pc group with 4 generators>
gap> NumberElements := true;;
gap> MultTable(S);
```

*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
e	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
f4	2	1	4	3	6	5	8	7	11	12	9	10	15	16	13	14
f3	3	4	1	2	7	8	5	6	10	9	12	11	14	13	16	15
f3*f4	4	3	2	1	8	7	6	5	12	11	10	9	16	15	14	13
f2	5	6	7	8	1	2	3	4	13	14	15	16	9	10	11	12
f2*f4	6	5	8	7	2	1	4	3	15	16	13	14	11	12	9	10
f2*f3	7	8	5	6	3	4	1	2	14	13	16	15	10	9	12	11
f2*f3*f4	8	7	6	5	4	3	2	1	16	15	14	13	12	11	10	9
f1	9	10	11	12	13	14	15	16	5	6	7	8	1	2	3	4
f1*f4	10	9	12	11	14	13	16	15	7	8	5	6	3	4	1	2
f1*f3	11	12	9	10	15	16	13	14	6	5	8	7	2	1	4	3
f1*f3*f4	12	11	10	9	16	15	14	13	8	7	6	5	4	3	2	1
f1*f2	13	14	15	16	9	10	11	12	1	2	3	4	5	6	7	8
f1*f2*f4	14	13	16	15	10	9	12	11	3	4	1	2	7	8	5	6
f1*f2*f3	15	16	13	14	11	12	9	10	2	1	4	3	6	5	8	7
f1*f2*f3*f4	16	15	14	13	12	11	10	9	4	3	2	1	8	7	6	5

Before the `SemidirectProduct` command will work, we *must* calculate the size of the first group, `Z4` in this case. Finding the size of the group establishes the elements of the group. Other commands that list the elements of the first group would also work, such as the `List` or `MultTable` commands.

GAP defines this non-abelian group of order 16 using four different generators `f1`, `f2`, `f3`, and `f4`. But if we look carefully, we see that `f1` generates a copy of Z_4, while `f3` and `f4` generate a copy of Z_8^*. It appears that the semi-direct product, like the direct product, contains copies of the two original groups within the product.

LEMMA 6.7

Let $G = H \ltimes_{\phi} K$ be the semi-direct product of K with H through the homomorphism ϕ. Suppose that e_1 is the identity element of H, and e_2 is the identity element of K. Then

$$\overline{H} = \{(h, e_2) \quad | \quad h \in H\}$$

is a subgroup of G, and

$$\overline{K} = \{(e_1, k) \quad | \quad k \in K\}$$

is a normal subgroup of G. Furthermore, $\overline{H} \approx H$, $\overline{K} \approx K$, and $\overline{H} \cap \overline{K}$ is the identity element of G.

PROOF We will use proposition 2.2 and observe that

$$(h, e_2)^{-1} = (h^{-1}, \phi_{h^{-1}}(e_2)) = (h^{-1}, e_2),$$

so

$$(h_1, e_2)\cdot(h_2, e_2)^{-1} = (h_1, e_2)\cdot(h_2^{-1}, e_2) = (h_1 \cdot h_2^{-1}, \phi_{h_2^{-1}}(e_2)\cdot e_2) = (h_1 \cdot h_2^{-1}, e_2).$$

Thus, whenever a and b are in $\overline{H}$, $a \cdot b^{-1}$ is in $\overline{H}$. So $\overline{H}$ is a subgroup.

Also,

$$(e_1, k)^{-1} = (e_1, \phi_{e_1}(k^{-1})) = (e_1, k^{-1}),$$

so

$$(e_1, k_1) \cdot (e_1, k_2)^{-1} = (e_1, k_1) \cdot (e_1, k_2^{-1}) = (e_1, \phi_{e_1}(k_1) \cdot k_2^{-1}) = (e_1, k_1 \cdot k_2^{-1}).$$

Thus, $a \cdot b^{-1}$ is in $\overline{K}$ whenever a and b are in $\overline{K}$. So $\overline{K}$ is also a subgroup by proposition 2.2. To show that this group is also a normal subgroup we look at

$$\begin{aligned}[(h, k_1) \cdot (e_1, k_2)] \cdot (h, k_1)^{-1} &= (h, \phi_{e_1}(k_1) \cdot k_2) \cdot (h^{-1}, \phi_{h^{-1}}(k_1^{-1})) \\ &= (e_1, \phi_{h^{-1}}(k_1 \cdot k_2) \cdot \phi_{h^{-1}}(k_1^{-1})) \\ &= (e_1, \phi_{h^{-1}}(k_1 \cdot k_2 \cdot k_1^{-1})).\end{aligned}$$

Since $g \cdot k \cdot g^{-1}$ is in $\overline{K}$ whenever k is in $\overline{K}$, by proposition 3.4 $\overline{K}$ is a normal subgroup of G.

Finally, the two mappings

$$f_1(h) = (h, e_2) \qquad \text{and} \quad f_2(k) = (e_1, k)$$

are isomorphisms from H onto $\overline{H}$ and K onto $\overline{K}$, respectively, as seen by the above computations. Also, it is clear that the intersections of the two groups give just $\{(e_1, e_2)\}$. ∎

Since the semi-direct product contains copies of the two smaller groups within itself, the natural question is whether an arbitrary group G can be expressed as a semi-direct product of two of its subgroups. The conditions for which this happens is set forth in the following theorem.

THEOREM 6.3: The Semi-Direct Product Theorem
Suppose that a group G has two subgroups H and N whose intersection is the identity element. Then if N is a normal subgroup of G and H is not a normal subgroup of $H \cdot N$, then there exists a nontrivial homomorphism ϕ from H to $\mathrm{Aut}(N)$ *such that*

$$H \cdot N \approx H \ltimes_{\phi} N.$$

PROOF Note that since H is a subgroup of G, and N is a normal subgroup we have by lemma 4.5 that $H \cdot N$ is a subgroup of G. We next want to define the homomorphism ϕ. For each h in H, we define

$$\phi_h(n) = h^{-1} \cdot n \cdot h$$

for all $n \in N$. We first need to show that ϕ_h is an automorphism on N for each h in H, and then we need to show that ϕ itself is a nontrivial homomorphism. Note that

$$\phi_h(n_1 \cdot n_2) = h^{-1} \cdot n_1 \cdot n_2 \cdot h = (h^{-1} \cdot n_1 \cdot h) \cdot (h^{-1} \cdot n_2 \cdot h) = \phi_h(n_1) \cdot \phi_h(n_2).$$

So ϕ_h is a homomorphism from N to N. Since

$$y \in \phi_h^{-1}(n) \Longleftrightarrow h^{-1} \cdot y \cdot h = n \Longleftrightarrow y = h \cdot n \cdot h^{-1}$$

we see that ϕ_h is a one-to-one and onto function. Thus, ϕ_h is an automorphism of N.

Next, we need to see that ϕ itself is a homomorphism from H to $\mathrm{Aut}(N)$. Note that

$$\begin{aligned}(\phi_{h_1} \cdot \phi_{h_2})(n) &= \phi_{h_2}(\phi_{h_1}(n)) \\ &= \phi_{h_2}(h_1^{-1} \cdot n \cdot h_1) \\ &= h_2^{-1} \cdot h_1^{-1} \cdot n \cdot h_1 \cdot h_2 \\ &= (h_1 \cdot h_2)^{-1} \cdot n \cdot (h_1 \cdot h_2) = \phi_{h_1 \cdot h_2}(n).\end{aligned}$$

So $\phi_{h_1} \cdot \phi_{h_2} = \phi_{(h_1 \cdot h_2)}$ and we see that ϕ is a homomorphism. In fact, the homomorphism must be nontrivial, because if $\phi_h(n) = n$ for all h and n, then since $\phi_h(n) = h^{-1} \cdot n \cdot h = n$ we have that $n \cdot h = h \cdot n$ for all h in H, and n in N. This would indicate that H is a normal subgroup of $H \cdot N$, which contradicts our original assumption. Thus, ϕ is a nontrivial homomorphism.

We can now proceed in a similar way that we proved the direct product theorem (6.1). However, it will be easier if we first show that every element in $H \cdot N$ can be *uniquely* written in the form $h \cdot n$, where $h \in H$ and $n \in N$.

Suppose that we have

$$h_1 \cdot n_1 = h_2 \cdot n_2.$$

Then $h_2^{-1} \cdot h_1 = n_2 \cdot n_1^{-1}$. Since this element is in both H and N, which has just the identity element in the intersection, we must have

$$h_2^{-1} \cdot h_1 = n_2 \cdot n_1^{-1} = e.$$

Therefore, $h_1 = h_2$ and $n_1 = n_2$. Thus, we have shown that every element of $H \cdot N$ is written uniquely as $h \cdot n$, where h is in H, and n is in N.

We now want to create a mapping

$$f : H \cdot N \to H \ltimes_{\phi} N$$

defined by

$$f(v) = (h, n),$$

where h and n are the unique elements such that $h \in H$, $n \in N$, and $v = h \cdot n$. The function f is one-to-one since the element (h, n) can only come from $h \cdot n$. Also, the element $h \cdot n$ maps to (h, n) so f is onto.

The final step is to show that f is a homomorphism. Let $v = h_1 \cdot n_1$, and $w = h_2 \cdot n_2$. Then

$$v \cdot w = h_1 \cdot n_1 \cdot h_2 \cdot n_2 = (h_1 \cdot h_2) \cdot (h_2^{-1} \cdot n_1 \cdot h_2 \cdot n_2).$$

Since N is a normal subgroup, $h_2^{-1} \cdot n_1 \cdot h_2$ is in N, and so $h_2^{-1} \cdot n_1 \cdot h_2 \cdot n_2$ is in N while $h_1 \cdot h_2$ is in H. Thus,

$$\begin{aligned} f(v \cdot w) &= f((h_1 \cdot h_2) \cdot (h_2^{-1} \cdot n_1 \cdot h_2 \cdot n_2)) \\ &= (h_1 \cdot h_2, h_2^{-1} \cdot n_1 \cdot h_2 \cdot n_2) \\ &= (h_1 \cdot h_2, \phi_{h_2}(n_1) \cdot n_2) \\ &= (h_1, n_1) \cdot (h_2, n_2) = f(v) \cdot f(w). \end{aligned}$$

So f is an isomorphism, and we have $H \cdot N \approx H \ltimes_{\phi} N$. □

Note that if both H and N are normal subgroups of $H \cdot N$, we have by corollary 6.1 that $H \cdot N \approx H \times N$.

We will use the semi-direct product theorem to define this product in *Mathematica*. After defining the two groups H and N using the same identity element, we must find the homomorphism ϕ from H to $\mathrm{Aut}(N)$. As in the case of the direct product, We will want to express every element of the form $h \cdot n$, where h is in H, and n is in N. From the definition, we see that

$$(h, e_2) \cdot (e_1, n) = (h \cdot e_1, \phi_{e_1}(e_2) \cdot n) = (h, n),$$

So for each generator a of H, and each generator b of N, we can calculate how **b . a** should be defined by evaluating $(e_1, b) \cdot (a, e_2) = (a, \phi_a(b))$. Thus we make a definition in *Mathematica* of the form

Define[b.a, a . $\phi_a(b)$ **]**

where we replace the expression $\phi_a(b)$ with its element of N.

Suppose we want to find a semi-direct product of Z_5 with Z_2.

InitGroup[e];
Define[a^2, e]
Define[1/a, a]
Z2 = Group[{a}]
Define[b^5, e]
Define[1/b, b^4]
Z5 = Group[{b}]

After loading the groups Z_2 and Z_5, we want to find a nontrivial homomorphism ϕ from Z_2 to $\mathrm{Aut}(Z_5)$. But $\mathrm{Aut}(Z_5) \approx Z_5^* \approx Z_4$. Since the element a is of order 2, ϕ_a must be of order 2 to keep the homomorphism from being trivial. But it is easy to find the one element of $\mathrm{Aut}(Z_5)$ of order 2:

$$\phi(n) = n^{-1}.$$

In fact, this will always be an automorphism whenever N is an abelian group. As long as N has an element that is not its own inverse, this automorphism will be of order 2. If we let $\phi_a(n) = n^{-1}$, then $\phi_a(b) = b^4$. Thus, the definition

Define[b.a, a.b.b.b.b]

completes the definition of the semi-direct product.

G = Group[{a, b}]
$\{e,\ a,\ b,\ a \cdot b,\ b \cdot b,\ a \cdot b \cdot b,\ b \cdot b \cdot b,\ a \cdot b \cdot b \cdot b,\ b \cdot b \cdot b \cdot b,\ a \cdot b \cdot b \cdot b \cdot b\}$

The corresponding GAP commands are

```
gap> f := FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g := f/[a^2, b^5, b*a/(a*b^4)];; a := g.1;; b := g.2;;
gap> L := ListGroup(g);
[ <identity...>, a, b, a*b, b^2, a*b^2, b^3, a*b^3, b^4, a*b^4 ]
gap> ResetTableOptions();
gap> MultTable(L);
*    |e     a     b     a*b   b^2   a*b^2 b^3   a*b^3 b^4   a*b^4
-----+-----------------------------------------------------------
e    |e     a     b     a*b   b^2   a*b^2 b^3   a*b^3 b^4   a*b^4
a    |a     e     a*b   b     a*b^2 b^2   a*b^3 b^3   a*b^4 b^4
b    |b     a*b^4 b^2   a     b^3   a*b   b^4   a*b^2 e     a*b^3
a*b  |a*b   b^4   a*b^2 e     a*b^3 b     a*b^4 b^2   a     b^3
b^2  |b^2   a*b^3 b^3   a*b^4 b^4   a     e     a*b   b     a*b^2
a*b^2|a*b^2 b^3   a*b^3 b^4   a*b^4 e     a     b     a*b   b^2
b^3  |b^3   a*b^2 b^4   a*b^3 e     a*b^4 b     a     b^2   a*b
a*b^3|a*b^3 b^2   a*b^4 b^3   a     b^4   a*b   e     a*b^2 b
b^4  |b^4   a*b   e     a*b^2 b     a*b^3 b^2   a*b^4 b^3   a
a*b^4|a*b^4 b     a     b^2   a*b   b^3   a*b^2 b^4   a*b^3 e
```

which show that this is a non-abelian group of order 10. If we ask GAP what this group is,

```
gap> StructureDescription(g);
"D10"
```

we find that this group is `D10`, which is GAP's way of saying the dihedral group that has 10 elements, or D_5.

DEFINITION 6.7 Let $n > 2$, and let ϕ be the homomorphism from $Z_2 = \{e, a\}$ to $\mathrm{Aut}(Z_n)$ given by

$$\phi_e(k) = k, \qquad \phi_a(k) = k^{-1}.$$

Then the semi-direct product $Z_2 \ltimes_\phi Z_n$ is called the *dihedral group of order* $2n$. It is denoted D_n, and is a non-abelian group of order $2n$.

The commands

InitGroup[e];
Define[a^2, e]
Define[b^*n*, e]
Define[1/a, a]
Define[1/b, b^(*n*–1)]
Define[b.a, a.(1/b)]
Dn = Group[{a, b}]

define the group D_n. The corresponding GAP commands are

```
gap> f := FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g := f/[a^2, b^n, a*b*a*b]; a := g.1;; b:= g.2;;
```

The symbol n must be replaced with an integer before executing these commands. When $n = 3$, we get a non-abelian group of order 6, so $D_3 \approx S_3$.

Note that the semi-direct product may greatly depend on the choice of the homomorphism ϕ. Consider finding the semi-direct products of Z_8 with Z_2. Since $\mathrm{Aut}(Z_8) \approx Z_8^*$ has three elements of order 2, there are three nontrivial homomorphisms from Z_2 to $\mathrm{Aut}(Z_8)$. One of these produces the dihedral group D_8 above, but the other two homomorphisms produce the groups

InitGroup[e];
Define[a^2, e]; Define[b^8, e]
Define[1/a, a]; Define[1/b, b^7]
Define[b.a, a.(b^3)]
G = Group[{a, b}]
and
InitGroup[e];
Define[a^2, e]; Define[b^8, e]
Define[1/a, a]; Define[1/b, b^7]
Define[b.a, a.(b^5)]
M = Group[{a, b}]

in *Mathematica*. These two groups along with D_8 can be entered in GAP at the same time as follows:

```
gap> f:= FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> D8:=f/[a^2, b^8, b*a/(a*b^7)];;
gap> G := f/[a^2, b^8, b*a/(a*b^3)];;
gap> M := f/[a^2, b^8, b*a/(a*b^5)];;
gap> StructureDescription(D8);
"D16"
gap> StructureDescription(G);
"QD16"
gap> StructureDescription(M);
"C8 : C2"
```

GAP's structure description shows that these three groups are all different. The group G is called a *quasidihedral group*, whereas the group M has no special name. GAP uses the colon to show a semi-direct of `C8` with `C2`. GAP uses "C8" for the cyclic group of order 8, that is, Z_8. Thus, structure description of M shows that it is a group of the form $Z_2 \ltimes Z_8$.

Here is another way of showing that the three groups are different:

```
gap> SetReducedMultiplication(D8);
gap> SetReducedMultiplication(G);
gap> SetReducedMultiplication(M);
gap> List(D8, x -> x^2);
[ <identity ...>, <identity ...>, b^2, b^-4, <identity ...>,
  <identity ...>, <identity ...>, <identity ...>, b^-2, b^2,
  b^-4, <identity ...>, <identity ...>, <identity ...>, b^-2,
  <identity ...> ]
gap> List(G, x -> x^2);
[ <identity ...>, <identity ...>, b^2, a^-1*b^-1*a^-1*b^-1,
  <identity ...>, a^-1*b^-1*a^-1*b^-1, <identity ...>,
  <identity ...>, b^-2, b^2, a^-1*b^-1*a^-1*b^-1,
  a^-1*b^-1*a^-1*b^-1, a^-1*b^-1*a^-1*b^-1, <identity ...>,
  b^-2, a^-1*b^-1*a^-1*b^-1 ]
gap> List(M, x -> x^2);
[ <identity ...>, <identity ...>, b^2, a^-1*b^-1*a^-1*b,
  <identity ...>, b^-2, a^-1*b^-1*a^-1*b, <identity ...>, b^-2,
  b^2, a^-1*b^-1*a^-1*b, b^2, b^-2, a^-1*b^-1*a^-1*b, b^-2, b^2]
```

This simple test shows that D_8 has nine elements of order 2, while the group G has five elements of order 2, and the group M has only three elements of order 2.

We see that the semi-direct product $Z_2 \ltimes_\phi Z_8$ depends on the choice of the homomorphism ϕ. In fact, even though the three elements of $\mathrm{Aut}(Z_8)$ of order 2 are essentially equivalent (since the automorphisms of Z_8^* included all permutations of these three elements), we see that the three elements produced three different semi-direct products.

This example is really more of an exception rather than a rule. Part of what makes this example unusual is that the automorphism group Z_8^* is abelian,

and hence does not have any nontrivial *inner* automorphisms. If two homomorphisms ϕ and f from H to $\mathrm{Aut}(N)$ are related through an inner automorphism of $\mathrm{Aut}(N)$, then the corresponding semi-direct products will if fact be isomorphic.

PROPOSITION 6.7

Let ϕ be a homomorphism from a group H to the group $\mathrm{Aut}(N)$. Suppose that f is another homomorphism such that

$$f_h(n) = w(\phi_h(w^{-1}(n))),$$

where $w(n)$ is an automorphism of N. Then $H \ltimes_f N \approx H \ltimes_\phi N$.

PROOF Let us write $G = H \ltimes_\phi M$, and $M = H \ltimes_f N$. These are two different groups, even though they are both written using ordered pairs. Let us define a mapping

$$v : G \to M$$

defined by

$$v((h, n)) = (h, w(n)).$$

Because $w(n)$ is one-to-one and onto, certainly v is one-to-one and onto. All we would have to check is that

$$v((h_1, n_1)) \cdot v((h_2, n_2)) = v((h_1, n_1) \cdot (h_2, n_2)).$$

We have that

$$\begin{aligned} v((h_1, n_1)) \cdot v((h_2, n_2)) &= (h_1, w(n_1)) \cdot (h_2, w(n_2)) \\ &= (h_1 \cdot h_2, f_{h_2}(w(n_1)) \cdot w(n_2)) \\ &= (h_1 \cdot h_2, w(\phi_{h_2}(w^{-1}(w(n_1)))) \cdot w(n_2)) \\ &= (h_1 \cdot h_2, w(\phi_{h_2}(n_1)) \cdot w(n_2)). \end{aligned}$$

On the other hand,

$$\begin{aligned} v((h_1, n_1) \cdot (h_2, n_2)) &= v((h_1 \cdot h_2, \phi_{h_2}(n_1) \cdot n_2)) \\ &= (h_1 \cdot h_2, w(\phi_{h_2}(n_1) \cdot n_2)) \\ &= (h_1 \cdot h_2, w(\phi_{h_2}(n_1)) \cdot w(n_2)). \end{aligned}$$

Since these are equal, we have an isomorphism. ∎

It is also clear that two homomorphisms ϕ and f are related through an automorphism of H, the semi-direct products must be isomorphic since we are merely relabeling the elements of H. As a result there will be many instances in which there will be only one non-isomorphic semi-direct product of K by H. In this case, we can denote *the* semi-direct product as $H \ltimes N$, without having to specify the homomorphism ϕ.

Problems for Chapter 6

Interactive Problems

6.1 Use GAP or *Mathematica* to define the group $Z_2 \times Z_6$. Show that this group is not isomorphic to Z_{12}.

6.2 Define the group $S_3 \times Z_2$ in *Mathematica* or GAP. Show that this group is not isomorphic to A_4.

Hint: Count elements of order 2.

6.3 Use *Mathematica*'s **PartitionsP** command or GAP's `NrPartitions` command to find the number of abelian groups of order 120,000.

For problems **6.4** through **6.7**: Find all of the automorphisms of the following groups.

Hint: For the non-abelian groups, find the inner automorphisms first.

6.4 S_3 **6.5** Z_{15}^* **6.6** D_4 **6.7** D_5

6.8 Show that there is only one semi-direct product $Z_2 \ltimes Z_8^*$. Which of the five groups of order 8 is this isomorphic to?

Hint: Use proposition 6.7.

6.9 Use *Mathematica* or GAP to find the only semi-direct product $Z_8^* \ltimes Z_8^*$. Is this group isomorphic to any of the three groups of order 16 found by considering $Z_2 \ltimes_\phi Z_8$?

6.10 Use *Mathematica* or GAP to define the only possible semi-direct product $Z_4 \ltimes Z_3$. Show that this group is different than both A_4 and $S_3 \times Z_2$.

6.11 From problems 6.1, 6.2, 6.10, and section 6.4, we have found six groups of order 12: Z_{12}, $Z_2 \times Z_6$, A_4, D_6, $S_3 \times Z_2$, and $Z_4 \ltimes Z_3$. Yet table 4.4 in chapter 4 indicates that there are only five non-isomorphic groups of order 12. Which two of these groups are isomorphic? Use *Mathematica* or GAP to show the isomorphism.

Non-Interactive Problems

6.12 We have shown by process of elimination that $Z_4 \times Z_2$ is isomorphic to Z_{15}^*. Demonstrate the isomorphism by giving multiplication tables for the two groups with the same pattern.

6.13 Demonstrate that $Z_3 \times Z_2$ is isomorphic to Z_6.

6.14 Construct a multiplication table for $Z_2 \times Z_8^*$.

6.15 Construct a multiplication table for $Z_3 \times Z_8^*$.

6.16 Let $G = H \times K$, and define

$$\overline{H} = \{(h, e) \quad | \quad h \in H\}$$

and

$$\overline{K} = \{(e, k) \quad | \quad k \in K\}.$$

Prove that $G/\overline{H} \approx K$ and $G/\overline{K} \approx H$.

6.17 Let n be any integer greater than 1. Prove that $Z_n \times Z_n$ is not isomorphic to Z_{n^2}.

For problems **6.18** through **6.20**: Find, up to isomorphism, all abelian groups of the following orders:

6.18 $|G| = 32$ **6.19** $|G| = 210$ **6.20** $|G| = 200$

6.21 What is the smallest positive integer n for which there are exactly four non-isomorphic abelian groups of order n?

6.22 Calculate the number of elements of order 4 in the groups

$$Z_{16}, \qquad Z_8 \times Z_2, \qquad Z_4 \times Z_4, \quad \text{and} \quad Z_4 \times Z_2 \times Z_2.$$

6.23 How many elements of order 25 are in $Z_5 \times Z_{25}$? (Do not do this exercise by brute force.)

6.24 An abelian group G of order 256 has 1 element of order 1, 7 elements of order 2, 24 elements of order 4, 96 elements of order 8, and 128 elements of order 16. Determine up to isomorphism the group G as a direct product of cyclic groups.

Hint: Use lemma 6.5 to determine the value of the function

$$f(x) = \sum_{k=1}^{j} \text{Min}(n_k, x)$$

for $x = 1$, 2, 3, and 4. Then use lemma 6.6 to determine how many times Z_2, Z_4, Z_8, and Z_{16} appear in the decomposition.

6.25 If an abelian group G of order 40 has exactly three elements of order 2, determine up to isomorphism the group G.

6.26 Classify the integers n for which the only abelian groups of order n are cyclic.

6.27 Prove that if G is a finite group of order n, then $\mathrm{Aut}(G)$ is isomorphic to a subgroup of S_{n-1}.

6.28 Prove that any finite group of order greater than 2 has at least two automorphisms.

6.29 Prove that if G is not abelian, then $\mathrm{Aut}(G)$ is not cyclic.

6.30 Find $\mathrm{Aut}(\mathbb{Z})$.

6.31 Find two non-isomorphic groups G and M for which $\mathrm{Aut}(G) \approx \mathrm{Aut}(M)$.

6.32 Let $\phi : Z_8^* \to \mathrm{Aut}(Z_8^*)$ be defined as follows: $\phi_1(x) = \phi_3(x) = x$ for all x in Z_8^*. $\phi_5(1) = \phi_7(1) = 1$. $\phi_5(3) = \phi_7(3) = 5$. $\phi_5(5) = \phi_7(5) = 3$. $\phi_5(7) = \phi_7(7) = 7$. Compute the following in $Z_8^* \ltimes_\phi Z_8^*$: $(5,3) \cdot (3,5)$, $(3,5) \cdot (5,3)$, $(7,5)^{-1}$.

6.33 Show that there is only one semi-direct product of the form $Z_3 \ltimes Z_8^*$. Form a multiplication table of this group. You have seen this group before. Do you recognize it?

6.34 Show that there is only one semi-direct product of the form $Z_2 \ltimes \mathbb{Z}$. Describe this group.

6.35 Show that there is only one semi-direct product of the form $\mathbb{Z} \ltimes \mathbb{Z}$. Describe this group.

6.36 Let G be any group, and let i be the identity mapping from $\mathrm{Aut}(G)$ to itself. We can define the semi-direct product $H = \mathrm{Aut}(G) \ltimes_i G$. The group H is called the *holomorph* of G. Show that every automorphism of G is the restriction of some inner automorphism of the holomorph H.

This page intentionally left blank

Chapter 7

The Search for Normal Subgroups

7.1 The Center of a Group

We saw several instances in the last chapter in which the structure of a group hinges on its normal subgroups. Thus, we will want to develop techniques for finding *all* of the normal subgroups of a given group G. We will discover in the process that some of the normal groups have additional properties. We will naturally concentrate our attention to non-abelian groups, since every subgroup of an abelian group is normal.

Let us begin by considering the quaternion group Q. This can be created in GAP by the command `InitQuaternions()`.

```
gap> InitQuaternions();
#I  default 'IsGeneratorsOfMagmaWithInverses' method returns
'true' for [ i, j ]
gap> MultTable(Q);
*     |(-1)*e (-1)*i (-1)*j (-1)*k k      j      i      e
------+-------------------------------------------------------
(-1)*e|e      i      j      k      (-1)*k (-1)*j (-1)*i (-1)*e
(-1)*i|i      (-1)*e k      (-1)*j j      (-1)*k e      (-1)*i
(-1)*j|j      (-1)*k (-1)*e i      (-1)*i e      k      (-1)*j
(-1)*k|k      j      (-1)*i (-1)*e e      i      (-1)*j (-1)*k
k     |(-1)*k (-1)*j i      e      (-1)*e (-1)*i j      k
j     |(-1)*j k      e      (-1)*i i      (-1)*e (-1)*k j
i     |(-1)*i e      (-1)*k j      (-1)*j k      (-1)*e i
e     |(-1)*e (-1)*i (-1)*j (-1)*k k      j      i      e
```

The equivalent in *Mathematica*® would be

```
InitGroup[e];
Define[i^4, e]; Define[j^2, i^2]
Define[j.i, i.i.i.j]
Define[1/i, i^3]; Define[1/j, i.i.j]
Q = Group[{i, j}]
MultTable[Q];
```

which produces table 7.1.

There is only one element of order 2 in this group, namely `(-1)*e` (or i^2 in *Mathematica*.) But this element has another important property. Notice that

TABLE 7.1: Multiplication table for Q

$\cdot$	e	i	j	i^2	$i \cdot j$	i^3	$i^2 \cdot j$	$i^3 \cdot j$
e	e	i	j	i^2	$i \cdot j$	i^3	$i^2 \cdot j$	$i^3 \cdot j$
i	i	i^2	$i \cdot j$	i^3	$i^2 \cdot j$	e	$i^3 \cdot j$	j
j	j	$i^3 \cdot j$	i^2	$i^2 \cdot j$	i	$i \cdot j$	e	i^3
i^2	i^2	i^3	$i^2 \cdot j$	e	$i^3 \cdot j$	i	j	$i \cdot j$
$i \cdot j$	$i \cdot j$	j	i^3	$i^3 \cdot j$	i^2	$i^2 \cdot j$	i	e
i^3	i^3	e	$i^3 \cdot j$	i	j	i^2	$i \cdot j$	$i^2 \cdot j$
$i^2 \cdot j$	$i^2 \cdot j$	$i \cdot j$	e	j	i^3	$i^3 \cdot j$	i^2	i
$i^3 \cdot j$	$i^3 \cdot j$	$i^2 \cdot j$	i	$i \cdot j$	e	j	i^3	i^2

the locations of the i^2 in table 7.1 form a symmetrical pattern along the main diagonal. This indicates that whenever $a \cdot b = i^2$, then $b \cdot a = i^2$ in Q. Hence $b = a^{-1} \cdot i^2 = i^2 \cdot a^{-1}$. Therefore, i^2 commutes with all of the elements of Q.

DEFINITION 7.1 Given a group G, the *center* of G is defined to be the set of elements x for which $x \cdot y = y \cdot x$ for all elements $y \in G$. The center of a group G is customarily denoted $Z(G)$ because of the German word for center, *zentrum*. [1, p. 150]

From this definition, we see that $i^2 \in Z(Q)$. It is also clear that $e \in Z(G)$ for all groups, since $e \cdot y = y \cdot e$. By examining table 7.1 we find that there are no other elements of Q in $Z(Q)$, so $Z(Q) = \{e, i^2\}$. This is obviously a subgroup, but it turns out to be a normal subgroup because of the following proposition.

PROPOSITION 7.1
Given a group G, then $Z(G)$ is a normal subgroup of G.

PROOF First, we need to show that $Z(G)$ is a subgroup of G. If x and y are in $Z(G)$, and a is any element in G, then

$$x \cdot y \cdot a = x \cdot a \cdot y = a \cdot x \cdot y.$$

So $x \cdot y$ commutes with all of the elements of G. Thus, $x \cdot y$ is in $Z(G)$.

Also, we have

$$x^{-1} \cdot a = (a^{-1} \cdot x)^{-1} = (x \cdot a^{-1})^{-1} = a \cdot x^{-1},$$

So x^{-1} must also be in $Z(G)$. Thus, by proposition 2.2, $Z(G)$ is a subgroup of G.

Next, we can see that

$$a \cdot x \cdot a^{-1} = x \cdot a \cdot a^{-1} = x.$$

So $a \cdot x \cdot a^{-1}$ is in $Z(G)$ whenever x is in $Z(G)$ and a is in G. Thus, by proposition 3.4, $Z(G)$ is a normal subgroup of G. □

We use the command **GroupCenter** to find the center of a group in *Mathematica*. For example, the command

Z = GroupCenter[Q]

verifies our earlier observation that $Z(Q) = \{e, i^2\}$. In GAP, the command is simply `Center` or `Centre`.

```
gap> List(Center(Q));
[ (-1)*e, e ]
```

Although the center always produces a normal subgroup, this subgroup is not always interesting. For example, *Mathematica* or GAP can show that the center of the group S_3 is just the identity element.

```
gap> S3 := Group( (1,2), (1,2,3) );
Group([ (1,2), (1,2,3) ])
gap> List(Center(S3));
[ () ]
```

Whenever the center is just the identity element, we say the group is *centerless.* In fact, all of the permutation groups S_n bigger than S_3 are centerless. Since the proof involves an even permutation, we will find the center of A_n at the same time.

PROPOSITION 7.2
If $n > 3$, then the groups S_n and A_n are centerless.

PROOF Suppose that ϕ is an element of S_n or A_n which is not the identity. We need to show that ϕ cannot be in the center of either S_n or A_n, which amounts to finding an element of A_n that does not commute with ϕ.

Since ϕ is not the identity, there is some number x that is not fixed by ϕ, say x is mapped to y. Since $n > 3$, there is at least one number not in the list $\{x, y, \phi(y)\}$. Let z be one of these remaining numbers. Finally, we let f be the 3-cycle (xyz).

Since f is an even permutation f is in A_n. Then $\phi \cdot f$ sends x to z, but $f \cdot \phi$ sends x to $\phi(y) \neq z$. Thus, $f \cdot \phi \neq \phi \cdot f$, and ϕ is not in the center of either A_n or S_n. □

The other extreme is if $Z(G)$ is the entire group G. This happens if, and only if, the group G is abelian.

Since $Z(N)$ is a normal subgroup of G, what is the quotient group? The answer is rather interesting.

PROPOSITION 7.3
If G is a group, then $G/Z(G) \approx \text{Inn}(G)$.

PROOF We begin by observing that the mapping

$$\phi : G \to \text{Inn}(G)$$

given by

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

is a homomorphism, as we saw in the proof of the semi-direct product theorem (6.3). By the definition of the inner automorphisms, this mapping is surjective. However, this mapping is not necessarily injective. Let us determine the kernel of ϕ.

Suppose that ϕ_x is the identity homomorphism. Then $\phi_x(y) = y$ for all y in G. This means that $x \cdot y \cdot x^{-1} = y$, or $x \cdot y = y \cdot x$, for all y in G. Thus, x is in the center of G.

Now, suppose x is in $Z(G)$. Then $\phi_x(y) = x \cdot y \cdot x^{-1} = y \cdot x \cdot x^{-1} = y$, so ϕ_x is the identity homomorphism. Thus the kernel of ϕ is precisely the center of $Z(G)$. Therefore, by the first isomorphism theorem (4.1), we have

$$G/Z(G) \approx \text{Inn}(G). \qquad \square$$

The center of a group possesses a characteristic that is even stronger than that of a normal subgroup. To illustrate this characteristic, consider the next proposition.

PROPOSITION 7.4
Let N be a normal subgroup of a group G. Then $Z(N)$ is a normal subgroup not only of N, but also of G.

PROOF Let g be an element of G, and z an element of $Z(N)$. We need to show that $g \cdot z \cdot g^{-1}$ is in $Z(N)$. Since N is a normal subgroup of G, we certainly know that $g \cdot z \cdot g^{-1}$ is in N, so the way to test that it is in $Z(N)$ is to show that it commutes with every element of N.

Let n be an element of N. We want to show that $g \cdot z \cdot g^{-1} \cdot n = n \cdot g \cdot z \cdot g^{-1}$. Let $h = g^{-1} \cdot n \cdot g$. Then h is in N, since N is normal in G. Also, $n = g \cdot h \cdot g^{-1}$, so

$$\begin{aligned} g \cdot z \cdot g^{-1} \cdot n &= (g \cdot z \cdot g^{-1}) \cdot (g \cdot h \cdot g^{-1}) = g \cdot z \cdot h \cdot g^{-1} = g \cdot h \cdot z \cdot g^{-1} \\ &= (g \cdot h \cdot g^{-1}) \cdot (g \cdot z \cdot g^{-1}) = n \cdot g \cdot z \cdot g^{-1}. \end{aligned}$$

Hence, $g \cdot z \cdot g^{-1}$ commutes with every element n in N, so $g \cdot z \cdot g^{-1}$ is in $Z(N)$. By proposition 3.4, we have that $Z(N)$ is a normal subgroup of G. □

This proposition demonstrates a rather unusual property of a center of a group. In general, the normal subgroup of a normal subgroup is not necessarily a normal subgroup. Consider $M = \{(), (12)(34), (13)(24), (14)(23)\}$, which is a normal subgroup of S_4, and $H = \{(), (12)(34)\}$, which is a normal subgroup of M.

```
gap> S4 := Group( (1,2),(1,2,3), (1,2,3,4) );
Group([ (1,2), (1,2,3), (1,2,3,4) ])
gap> M := Group( (1,2)(3,4), (1,3)(2,4) );
Group([ (1,2)(3,4), (1,3)(2,4) ])
gap> H := Group( (1,2)(3,4) );
Group([ (1,2)(3,4) ])
gap> IsNormal(S4,M);
true
gap> IsNormal(M,H);
true
gap> IsNormal(S4,H);
false
```

So H is not a normal subgroup of S_4.

However, the center of a group $Z(N)$ is a normal subgroup of G, even though $Z(N)$ contains no information about the larger group G. Any group that contains N as a normal subgroup, such as a semi-direct product of N by another group, will have $Z(N)$ as a normal subgroup.

7.2 The Normalizer and Normal Closure Subgroups

In the last section, we found a subgroup of N that was not only normal, but also was normal in any group G for which N was a normal subgroup. In this section, we will essentially turn the question around: Given a subgroup H of G, can we find a subgroup N of G for which H lies inside of N as a normal subgroup?

DEFINITION 7.2 Let S be a *subset* of a group G. We define the *normalizer of S by G,* denoted $N_G(S)$, to be the set

$$N_G(S) = \{g \in G \quad | \quad g \cdot S \cdot g^{-1} = S\}.$$

Notice that this definition allows for S to be merely a *subset* of G, not necessarily a subgroup. We will later find uses for having a more generalized definition. For now, let us show that the normalizer has some of the properties that we are looking for.

PROPOSITION 7.5
Let S be a subset of the group G. Then $N_G(S)$ is a subgroup of G.

PROOF Suppose x and y are in $N_G(S)$. Then $x \cdot S \cdot x^{-1} = S$, and $y \cdot S \cdot y^{-1} = S$. Thus, $S = y^{-1} \cdot S \cdot y$, and so

$$(x \cdot y^{-1}) \cdot S \cdot (x \cdot y^{-1})^{-1} = x \cdot (y^{-1} \cdot S \cdot y) \cdot x^{-1} = x \cdot S \cdot x^{-1} = S.$$

Thus, $x \cdot y^{-1}$ is in $N_G(S)$, and so by proposition 2.2, $N_G(S)$ is a subgroup of G. □

If, in addition, S is a subgroup of G, then the normalizer lives up to its name.

PROPOSITION 7.6
Let H be a subgroup of the group G. Then $N_G(H)$ is the largest subgroup of G that contains H as a normal subgroup.

PROOF First, we must check to see that H is a normal subgroup of $N_G(H)$. But this is obvious, since $g \cdot H \cdot g^{-1} = H$ for all g in $N_G(H)$.

Next, we must see that $N_G(H)$ is the largest such group. Suppose that Y is another subgroup of G that contained H as a normal subgroup. Then $y \cdot H \cdot y^{-1} = H$ for all $y \in Y$. Thus, $Y \subseteq N_G(H)$.

Since any subgroup of G that contains H as a normal subgroup is itself contained in $N_G(H)$, we have that $N_G(H)$ is the largest such group. □

The *Mathematica* command

Normalizer[G, H]

finds the normalizer $N_G(H)$ of the set H in G. Suppose we consider the quaternion group Q.

```
InitGroup[e];
Define[i^4, e]; Define[j^2,i^2]
Define[j.i, i.i.i.j]
Define[1/i, i^3]; Define[1/j, i.i.j]
Q = Group[{i, j}]
```

Let begin by finding the normalizer of a single element i. The *Mathematica* command

```
H = Normalizer[Q, {i}]
```

gives the subgroup of order 4 generated by i, namely $\{e, i, i^2, i^3\}$. We could now consider the normalizer of this subgroup by Q.

Normalizer[Q, H]

This gives us the entire group Q, the largest subgroup of Q for which H is normal. In general, whenever H is a normal subgroup of G, the normalizer of H by G will be the whole group G.

In GAP, we have two different commands to do what the *Mathematica* command **Normalizer** does. If we have just a single element, we use the `Centralizer` command to find $N_G(\{g\})$. When the GAP's `Normalizer` command is used with a single element, GAP finds the normalizer of the subgroup that is generated by this element, hence $N_G(H)$, for $H = [g]$.

```
gap> InitQuaternions();
#I  default 'IsGeneratorsOfMagmaWithInverses' method returns
'true' for [ i, j ]
gap> List(Centralizer(Q,i));
[ (-1)*e, (-1)*i, i, e ]
gap> List(Normalizer(Q,i));
[ (-1)*e, (-1)*i, (-1)*j, (-1)*k, k, j, i, e ]
```

This points out that $N_G(\{g\})$ is not the same thing as $N_G([g])$, the normalizer of the group generated by g.

In *Mathematica*, we can find the normalizer of any subset, even one that is not a subgroup. For example, the normalizer of the subset $\{i, j\}$ is

Normalizer[Q,{i, j}]
$\{e, i \cdot i\}$

which contains neither i nor j. Only when H is a *subgroup* or a *single element* can we be assured that $N_G(H)$ will contain H. In the latter case, when H is a single element g, $N_G(\{g\})$ will consist of all elements of G that commute with g.

We have seen that the normalizer of a subgroup H by G finds the largest subgroup of G that contains H as a normal subgroup. What if we asked for the *smallest* subgroup containing H that is a normal subgroup of G? Whether H is a subgroup or a subset, we can use the following proposition.

PROPOSITION 7.7
Let S be a subset *of a group G. Then the smallest group containing S that is a normal subgroup of G is given by*

$$N^* = \bigcap_{N \in L} N,$$

where L denotes the collection of normal subgroups of G that contain S.

PROOF The group G itself is in the collection L, so this collection is not empty. Thus, by proposition 2.3, N^* is a subgroup of G.

Also, since each N in the collection contained the set S, the intersection will also contain S. All that needs to be shown is that N^* is normal.

If n is an element of N^*, and g is an element of G, then since each N is a normal subgroup of G, and n would be in all of the groups N,

$$g \cdot n \cdot g^{-1} \in N \quad \text{for all} \quad N \in L.$$

Thus, $g \cdot n \cdot g^{-1}$ is in the intersection of all of the N's, which is N^*. Hence, by proposition 3.4, N^* is a normal subgroup of G. □

We will call this subgroup the *normal closure* of S. The *Mathematica* command

NormalClosure[G, S]

computes this subgroup for the subset S. In GAP, S must be a *subgroup* for this to work. So ironically, we first have to find the subgroup generated by a set before finding the normal closure. Thus, for a single element, we use

```
gap> List(NormalClosure(Q, Group(i) ) );
[ (-1)*e, (-1)*i, i, e ]
```

With this command we can systematically find *all* normal subgroups of a given group. For example, suppose we want to find all of the normal subgroups of S_3, using the generators a and b. We would like to see if there are any other normal subgroups besides the two trivial groups. Since a proper subgroup must contain one of the elements $\{a, b, a \cdot b, b^2, a \cdot b^2\}$, we have five groups to try.

```
gap> f:=FreeGroup("a","b");; a:=f.1;; b:=f.2;;
gap> S3:=f/[a^2,b^3,b*a*b*a];; a:=S3.1;; b:=S3.2;;
gap> List(NormalClosure(S3,Group(a)));
[ <identity ...>, a, b, a*b, a*b*a, b*a ]
gap> List(NormalClosure(S3,Group(b)));
[ <identity ...>, b, b^2 ]
gap> List(NormalClosure(S3,Group(a*b)));
[ <identity ...>, a, b, a*b, a*b*a, b*a ]
gap> List(NormalClosure(S3,Group(b^2)));
[ <identity ...>, b, b^2 ]
gap> List(NormalClosure(S3,Group(a*b^2)));
[ <identity ...>, a, b, a*b, a*b*a, b*a ]
```

We see that using b and b^2 produces the normal subgroup of order 3, A_3. The other elements produced the whole group. In fact, if we considered a normal subgroup generated by *two* elements, it is obvious that this would have to contain a normal subgroup already found. But the smallest found was A_3, and no larger subgroup could still be proper. Thus, we have used GAP to *prove* that the only proper normal subgroup of S_3 is A_3. Similar commands will also work in *Mathematica*.

This method of exhaustion works well for small groups, but one can imagine that this method would be time consuming for larger groups. In the next section, we will find a shortcut so that we will not have to try every element of the group, but rather just a handful of elements.

7.3 Conjugacy Classes and Simple Groups

In the last section, we used the GAP command `NormalClosure(G, S)` to find the smallest group containing the subset S that was a normal group of G. Let us look closely at how this command works. We know that if the element a is in this normal group, then $g^{-1} \cdot a \cdot g$ must also be in the group for all g in G. Many of the elements that must be in the normal subgroup can be found in this way.

DEFINITION 7.3 Let G be a group. We say that the element u is *conjugate* to the element v if there exists an element g in G such that $u = g^{-1} \cdot v \cdot g$.

Note that every element is conjugate to itself, for we can let g be the identity element. Also note that if u is conjugate to v, then v is also conjugate to u. Finally, if u is conjugate to v, and v in turn is conjugate to w, we can see that u is conjugate to w. This is easy to see, since there is a g and h such that $u = g^{-1} \cdot v \cdot g$ and $v = h^{-1} \cdot w \cdot h$. Then

$$u = g^{-1} \cdot v \cdot g = g^{-1} \cdot (h^{-1} \cdot w \cdot h) \cdot g = (h \cdot g)^{-1} \cdot w \cdot (h \cdot g).$$

Recall that in definition 3.5, we defined an equivalence relationship as any relationship having three properties:

1. Every element u is equivalent to itself.
2. If u is equivalent to v, then v is equivalent to u.
3. If u is equivalent to v, and v in turn is equivalent to w, then u is equivalent to w.

These were called the reflexive, symmetric, and transitive properties. We used the equivalence relationships of cosets in section 3.4 to form a partition of the group, which gave us the quotient groups. In the same way, we can use the equivalence relationship of conjugates to form a different partition of the group, called *conjugacy classes*. Unlike cosets, though, the conjugacy classes will not be all the same size. The conjugacy class containing the element u is given by

$$\{g^{-1} \cdot u \cdot g \quad | \quad g \in G\}$$

The command for finding all of the conjugacy classes of a group G for both *Mathematica* and GAP is `ConjugacyClasses`. Let us find the conjugacy classes of S_4, which are generated by the cycles (1 2) and (2 3 4).

```
gap> S4 := Group( (1,2), (2,3,4) );
Group([ (1,2), (2,3,4) ])
gap> L := ConjugacyClasses(S4);
[ ()^G, (1,2)^G, (1,2)(3,4)^G, (1,2,3)^G, (1,2,3,4)^G ]
```

GAP lists five conjugacy classes, but abbreviates each in the form `x^G`. Recall that GAP uses $x\hat{}y$ to represent $y^{-1}xy$, so it makes sense that `x^G` would mean $\{g^{-1} \cdot x \cdot g \quad | \quad g \in G\}$. Yet one must use the command

```
gap> ConjugacyClass(S4, (1,2));
(1,2)^G
```

to enter a particular conjugacy class into GAP. To see all of the elements in each conjugacy class, we can use a nested `List` command.

```
gap> List(L, x -> List(x));
[ [ () ], [ (1,2), (1,3), (1,4), (2,3), (2,4), (3,4) ],
  [ (1,2)(3,4), (1,3)(2,4), (1,4)(2,3) ],
  [ (1,2,3), (1,3,2), (1,4,2), (1,2,4), (1,3,4), (1,4,3),
    (2,4,3), (2,3,4) ],
  [ (1,2,3,4), (1,2,4,3), (1,3,2,4), (1,3,4,2), (1,4,2,3),
    (1,4,3,2) ] ]
```

The corresponding *Mathematica* commands for listing these conjugacy classes, albeit in a different order, are as follows:

S4 = Group[{**C**[**1,2**], **C**[**2,3,4**]}]
ConjugacyClasses[**S4**]

The identity element is in a class by itself since $g^{-1} \cdot e \cdot g$ will always produce e. But the cycle notation reveals an interesting fact about the other four classes: one contains all of the transpositions, one contains all of the 3-cycles, one contains all of the 4-cycles, and one conjugacy class contains the products of two disjoint transpositions. Problems 5.36 and 5.37 may help shed some light on why this happens.

The conjugacy classes are very useful for finding normal subgroups, since whenever one element of a conjugacy class is in a normal subgroup of G, the entire conjugacy class must be in the normal subgroup. Thus, in order to find *all* normal subgroups of S_4 we only have to try the different combinations of the conjugacy classes. Furthermore, the identity element is guaranteed to be in every subgroup. So to find all of the nontrivial normal subgroups, we only have to consider using one element from each conjugacy class besides the identity. Using GAP's list of the conjugacy classes shows that it selects the elements

$$S = \{(1,2),\ (1,2)(3,4),\ (1,2,3),\ (1,2,3,4)\}.$$

So we can consider taking the normal closure of any nontrivial subset of S. Thus, any nontrivial normal subgroup of S_4 must be one of the 14 possible groups:

NormalClosure[S4, { C[1,2] }]
NormalClosure[S4, { C[1,2].C[3,4] }]
NormalClosure[S4, { C[1,2,3] }]
NormalClosure[S4, { C[1,2,3,4] }]
NormalClosure[S4, { C[1,2] , C[1,2].C[3,4] }]
NormalClosure[S4, { C[1,2] , C[1,2,3] }]
NormalClosure[S4, { C[1,2] , C[1,2,3,4] }]
NormalClosure[S4, { C[1,2].C[3,4] , C[1,2,3] }]
NormalClosure[S4, { C[1,2].C[3,4] , C[1,2,3,4] }]
NormalClosure[S4, { C[1,2,3] , C[1,2,3,4] }]
NormalClosure[S4, { C[1,2] , C[1,2].C[3,4], C[1,2,3] }]
NormalClosure[S4, { C[1,2] , C[1,2].C[3,4], C[1,2,3,4] }]
NormalClosure[S4, { C[1,2] , C[1,2,3] , C[1,2,3,4] }]
NormalClosure[S4, { C[1,2].C[3,4], C[1,2,3] , C[1,2,3,4] }]

The 15th combination

NormalClosure[S4,{C[1,2], C[1,2,3], C[1,2,3,4], C[1,2].C[3,4]}]

obviously would give us the whole group. We can try these out in GAP as follows:

```
gap> Size(NormalClosure(S4,Group( (1,2) ) ) );
24
gap> Size(NormalClosure(S4,Group( (1,2)(3,4) ) ) );
4
gap> List(NormalClosure(S4,Group( (1,2)(3,4) ) ) );
[ (), (1,2)(3,4), (1,3)(2,4), (1,4)(2,3) ]
gap> Size(NormalClosure(S4,Group( (1,2,3) ) ) );
12
gap> List(NormalClosure(S4,Group( (1,2,3) ) ) );
[ (), (1,3,2), (1,2,3), (1,4,3), (2,4,3), (1,3)(2,4), (1,2,4),
  (1,4)(2,3), (2,3,4), (1,3,4), (1,2)(3,4), (1,4,2) ]
gap> Size(NormalClosure(S4,Group( (1,2,3,4) ) ) );
24
```

Although this only does 4 of the 14 combinations, with a little logic we see that all other combinations will produce one of the groups we see here. If either (1 2) or (1 2 3 4) is included, we would have all 24 elements. If (1 2 3) is included, then we might as well include (1 2)(3 4), since this was in the normal subgroup. Note that lemma 5.2 predicts that the normal closure of (1 2) is S_4, and the normal closure of (1 2 3) is A_4 as guaranteed by proposition 5.1. The normal closure of (1 2)(3 4) produces a normal subgroup of order 4 isomorphic to Z_8^*. Thus, by using the conjugacy classes we have found that the only proper normal subgroups of S_4 are A_4 and the group isomorphic to Z_8^*.

If we repeat this procedure with the group A_5 (which also has only five conjugacy classes), GAP or *Mathematica* shows that there are no proper normal subgroups of A_5. (See problem 7.19 for a non-computerized way to prove this.)

```
gap> A5 := Group( (1,2,3), (3,4,5) );
Group([ (1,2,3), (3,4,5) ])
gap> ConjugacyClasses(A5);
[ ()^G, (1,2)(3,4)^G, (1,2,3)^G, (1,2,3,4,5)^G, (1,2,3,5,4)^G ]
gap> Size(NormalClosure(A5, Group( (1,2)(3,4) ) ) );
60
gap> Size(NormalClosure(A5, Group( (1,2,3) ) ) );
60
gap> Size(NormalClosure(A5, Group( (1,2,3,4,5) ) ) );
60
gap> Size(NormalClosure(A5, Group( (1,2,3,5,4) ) ) );
60
```

Since the normal closure of any of these four elements yields the whole group, there can be no nontrivial normal subgroups of A_5.

DEFINITION 7.4 A group is said to be *simple* if it contains no normal subgroups besides itself and the identity subgroup.

The groups Z_p, for p a prime number, are the first examples we have seen of simple groups. We now have seen an example of a non-cyclic simple group, A_5. In fact this is the *smallest* non-cyclic simple group! (See problem 7.39.) GAP can prove that the group is simple in one step.

```
gap> IsSimple(A5);
true
```

Let us find other simple groups. The natural place to look is higher order alternating groups. We begin by showing that all 3-cycles are in one conjugacy class.

LEMMA 7.1
If $n > 4$, any two 3-cycles are conjugate in A_n. Furthermore, the conjugate of a 3-cycle is again a 3-cycle.

PROOF We begin by showing that the conjugate of a 3-cycle is again a 3-cycle. Let $(a\,b\,c)$ be a 3-cycle, and let ϕ be any permutation in A_n. Define the values $x = \phi(a)$, $y = \phi(b)$, and $z = \phi(c)$. Then we can compute

$$\phi^{-1} \cdot (a\,b\,c) \cdot \phi = (x\,y\,z).$$

Thus the conjugate of a 3-cycle is another 3-cycle.

Next we will show that any 3-cycle is conjugate to the element $(1\,2\,3)$ in A_n. Let $(u\,v\,w)$ be a 3-cycle. Since $n > 4$ there must be at least two numbers

not mentioned in this 3-cycle, so we will call two of them x and y. Consider the permutation

$$\phi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & \cdots \\ u & v & w & x & y & \cdots \end{pmatrix}.$$

Here, the dots indicate that when $n > 5$, we can complete the permutation in any way so that the numbers on the bottom row will be a permutation of the numbers 1 through n.

Now ϕ will either be an even permutation or an odd permutation. If ϕ is an odd permutation, we can consider instead the permutation

$$\phi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & \cdots \\ u & v & w & y & x & \cdots \end{pmatrix}.$$

So we may assume that ϕ is an even permutation. Thus ϕ is in A_n, and we can compute

$$\phi^{-1} \cdot (1\,2\,3) \cdot \phi = (u\,v\,w).$$

Therefore, any 3-cycle is conjugate to $(1\,2\,3)$, and so any two 3-cycles are conjugate to each other in A_n whenever $n > 4$. □

With this lemma, we can show that A_n will be a simple group whenever $n > 4$. This was originally proved by Abel using a long case-by-case argument. Since GAP or *Mathematica* has already shown that A_5 is simple, most of the cases can be covered at once.

THEOREM 7.1: Abel's Theorem
The alternating group A_n is simple for all $n > 4$.

PROOF Suppose that N is a proper normal subgroup of A_n, and let ϕ be an element of N besides the identity. By proposition 7.2, A_n is centerless. Since proposition 5.1 tells us that A_n is generated by 3-cycles, there must be at least one 3-cycle that does not commute with ϕ, say $(a\,b\,c)$. Thus, $\phi \cdot (a\,b\,c)$ is not equal to $(a\,b\,c) \cdot \phi$, or equivalently, $(a\,b\,c) \cdot \phi \cdot (a\,c\,b) \cdot \phi^{-1}$ is not the identity element.

Since N is a normal subgroup, $(a\,b\,c) \cdot \phi \cdot (a\,c\,b)$ must be in N. Therefore, $(a\,b\,c) \cdot \phi \cdot (a\,c\,b) \cdot \phi^{-1}$ must also be in N. But $\phi \cdot (a\,c\,b) \cdot \phi^{-1}$ is the conjugate of a 3-cycle, so by lemma 7.1 this is also a 3-cycle, say $(x\,y\,z)$. Thus, N contains a product of two 3-cycles, $(a\,b\,c) \cdot (x\,y\,z)$, which is not the identity.

Suppose that the cycles $(a\,b\,c)$ and $(x\,y\,z)$ are *disjoint* 3-cycles. Then we can conjugate the product by the 3-cycle $(c\,z\,y)$ to get another element in N:

$$(c\,y\,z) \cdot [(a\,b\,c) \cdot (x\,y\,z)] \cdot (c\,z\,y) = (a\,b\,z) \cdot (c\,y\,x).$$

We now have two elements of N that consist of two disjoint 3-cycles. If we multiply these two elements together we get

$$[(a\,b\,z) \cdot (c\,y\,x)] \cdot [(a\,b\,c) \cdot (x\,y\,z)] = (a\,c\,z\,b\,x) = (a\,c\,x) \cdot (x\,z\,b).$$

This must also be in N. Thus N contains a product of two 3-cycles that are *not* disjoint. In essence we can say that there is a non-identity element of N that moves at most five numbers, labeled a, b, c, x, and z.

Here's where we can take advantage of the fact that A_5 is known to be simple. Consider the subgroup H of A_n consisting of all even permutations of the five numbers a, b, c, x, and z. We have just showed that there is a nontrivial intersection of N and H. Let this intersection be M. Whenever x is in M and h is in H, then $h \cdot x \cdot h^{-1}$ is in both H and N. Thus $h \cdot x \cdot h^{-1}$ is in M. Hence M is a nontrivial normal subgroup of H.

But H is isomorphic to A_5 which we have proven using *Mathematica* or GAP to be a simple group. Thus M must be all of H. In particular M contains a 3-cycle, and so N contains a 3-cycle. By lemma 7.1 all 3-cycles of A_n are conjugate, so N contains all 3-cycles of A_n. Finally, by proposition 5.1 the 3-cycles generate A_n, so N must be all of A_n. Therefore, A_n is simple whenever $n > 4$. □

COROLLARY 7.1
If $n > 4$ then the only proper normal subgroup of S_n is A_n

PROOF Suppose that there were another normal subgroup, N. Then the intersection of N with A_n would be another normal subgroup of S_n, and so would be a normal subgroup of A_n. Since A_n is simple for $n > 4$, this intersection must either be the identity or all of A_n.

Suppose that the intersection is all of A_n. Then N contains A_n, and if N is not equal to A_n, N would contain more than half of the elements of S_n. But this would contradict Lagrange's theorem (3.1) unless $N = S_n$.

Suppose that the intersection of N and A_n is just the identity element. Then since both N and A_n are normal subgroups, we have by corollary 6.1,

$$N \cdot A_n \approx N \times A_n.$$

If N is not just the identity element, this quickly leads to a contradiction, for N could have order of at most 2, telling us that S_n was isomorphic to $Z_2 \times A_n$. But this is ridiculous, for we saw in proposition 7.2 that S_n was centerless, whereas $Z_2 \times A_n$ has both $(0, (\ ))$ and $(1, (\ ))$ in its center. Therefore, the only normal subgroups of S_n for $n > 4$ are S_n itself, A_n, and the identity element. □

We now have found two sequences of simple groups, namely Z_p for p being a prime number, and A_n for all $n > 4$. Are any of the other groups that we have looked at simple groups? Consider the group $\mathrm{Aut}(Z_{24}^*)$, a group of order 168 generated by the 149th and 735th permutation elements.

InitPermMultiplication
A = Group[{149, 735}]

As large as this group is, *Mathematica* can still quickly find the conjugacy classes.

ConjugacyClasses[A]
{{1}, {27, 61, 87, 122, 270, 404, 593, 640, 714, 735, 775,
1582, 1807, 2380, 2691, 3032, 3151, 3755, 4017, 4476, 4498},
{149, 187, 244, 357, 374, 467, 548, 558, 856, 1014, 1123, 1311,
1362, 1392, 1402, 1432, 1461, 1622, 1649, 1775, 1851, 1881, 2032,
2151, 2258, 2345, 2366, 2510, 2592, 2647, 2677, 2821, 2918, 3019,
3099, 3177, 3195, 3276, 3412, 3508, 3689, 3741, 3817, 3898, 3973,
3991, 4098, 4205, 4366, 4384, 4410, 4428, 4616, 4713, 4817, 4970},
{231, 331, 437, 496, 670, 684, 753, 793, 908, 1079, 1088, 1229,
1496, 1662, 1692, 1837, 1992, 2042, 2201, 2304, 2476, 2632,
2721, 2787, 2900, 3059, 3133, 3298, 3476, 3595, 3702, 3776,
3876, 4035, 4151, 4269, 4536, 4558, 4595, 4735, 4874, 4931},
{918, 970, 1185, 1267, 1475, 1796, 2002, 2069, 2240, 2471, 2562, 2761,
2981, 3336, 3372, 3573, 3622, 3958, 4156, 4309, 4581, 4753, 4904, 4965},
{953, 1052, 1133, 1202, 1537, 1732, 1962, 2107, 2183, 2418, 2602, 2847,
2963, 3358, 3455, 3486, 3662, 3936, 4190, 4226, 4657, 4695, 4847, 5023}}

So we have six conjugacy classes of this group, one of which is just the identity. The other five classes can be represented by first element in each list, which in *Mathematica* are the 27th, 149th, 231st, 918th, and 953rd permutations. To get this list in GAP, we can first define the group generated by the permutations $(1,2,3)(4,6,5)$ and $(2,4)(6,7)$.

```
gap> A := Group( (1,2,3)(4,6,5), (2,4)(6,7) );
Group([ (1,2,3)(4,6,5), (2,4)(6,7) ])
gap> L := ConjugacyClasses(A);
[ ()^G, (3,5)(6,7)^G, (2,3,4,5)(6,7)^G, (2,3,6)(4,5,7)^G,
  (1,2,3,4,6,7,5)^G, (1,2,3,5,7,4,6)^G ]
gap> List(L, x->Size(x));
[ 1, 21, 42, 56, 24, 24 ]
```

Once again, we see six conjugacy classes, one being the identity element, and the other five represented by the permutations $(3\,5)(6\,7)$, $(2\,3\,4\,5)(6\,7)$, $(2\,3\,6)(4\,5\,7)$, $(1\,2\,3\,4\,6\,7\,5)$, and $(1\,2\,3\,5\,7\,4\,6)$. We can then verify that the normal closure of each of these five elements yields the whole group.

```
gap> Size(NormalClosure(A, Group( (3,5)(6,7) ) ) );
168
gap> Size(NormalClosure(A, Group( (2,3,4,5)(6,7) ) ) );
168
gap> Size(NormalClosure(A, Group( (2,3,6)(4,5,7) ) ) );
168
gap> Size(NormalClosure(A, Group( (1,2,3,4,6,7,5) ) ) );
168
gap> Size(NormalClosure(A, Group( (1,2,3,5,7,4,6)  ) ) );
168
```

Thus, any proper normal subgroup cannot contain any of these five elements; we have shown that there are no proper normal subgroups, so $\mathrm{Aut}(Z_{24}^*)$ is a simple group. This is slightly easier in *Mathematica*:

NormalClosure[**A,** {**27**}]
NormalClosure[**A,** {**149**}]
NormalClosure[**A,** {**231**}]
NormalClosure[**A,** {**918**}]
NormalClosure[**A,** {**953**}]

This is the second largest non-cyclic simple group. (A_5 is the smallest and A_6 is the third smallest.) See problems 7.22 through 7.25 for more examples of simple groups.

In fact, $\mathrm{Aut}(Z_{24}^*)$ is the beginning of yet another infinite family of simple groups, called the Chevalley groups. We will not go into all of the ways this group can be generalized to produce these other groups, but we will mention an important result that has taken place during the 20th century. It was once thought that *all* finite simple groups were either the cyclic groups of prime order, the alternating groups, or one of the Chevalley or twisted Chevalley groups. (One of these groups turns out to be not quite simple. Yet taking half of the elements forms a new simple group, just as we took half of the elements of S_n to form the simple groups A_n.) But there were several other simple groups that were discovered, called *sporadic* groups. In the 1960s and 1970s it was proved that there are exactly 26 sporadic groups, ranging in size from a mere 7,920 elements to the monstrous 808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000 elements! These 26 sporadic groups are listed in [13]. Because these have been proven to be the only sporadic groups, all finite simple groups are now known.

7.4 The Class Equation and Sylow's Theorems

In working with the conjugacy classes from the last section, we may have noticed a pattern in the *size* of each of the conjugacy classes. For example, the conjugacy classes of S_4 are given by

```
gap> S4 := Group( (1,2), (2,3,4) );
Group([ (1,2), (2,3,4) ])
gap> L := ConjugacyClasses(S4);
[ ()^G, (1,2)^G, (1,2)(3,4)^G, (1,2,3)^G, (1,2,3,4)^G ]
gap> List(L, x -> List(x));
[ [ () ], [ (1,2), (1,3), (1,4), (2,3), (2,4), (3,4) ],
  [ (1,2)(3,4), (1,3)(2,4), (1,4)(2,3) ],
  [ (1,2,3), (1,3,2), (1,4,2), (1,2,4), (1,3,4), (1,4,3),
    (2,4,3), (2,3,4) ],
```

```
[ (1,2,3,4), (1,2,4,3), (1,3,2,4), (1,3,4,2), (1,4,2,3),
  (1,4,3,2) ] ]
```

The first class has only the identity element, the class with the transpositions has exactly six elements, while the other classes are of orders 3, 8, and 6. Immediately we see that the number of elements in the classes may be different. We have the obvious relationship

$$1 + 6 + 3 + 8 + 6 = 24,$$

the order of the group, since every element in the group belongs to one and only one conjugacy class. Is there another pattern? Let us compare this with the conjugacy classes of $\mathrm{Aut}(Z_{24}^*)$. There were six conjugacy classes of size 1, 21, 42, 56, 24, and 24. We can check that

$$1 + 21 + 42 + 56 + 24 + 24 = 168.$$

But another pattern is becoming clear that is akin to Lagrange's theorem (3.1). Notice that the number of elements in each class is always a *divisor of the order of the group.*

LEMMA 7.2

Let G be a finite group, and let g be an element of G. Then the number of elements of G that are conjugate to g is given by

$$\frac{|G|}{|N_G(\{g\})|},$$

where $N_G(\{g\})$ denotes the normalizer of the single element $\{g\}$.

PROOF We saw in proposition 7.5 that $N_G(\{g\})$ is a subgroup of G. We want to determine all possible conjugates of the element g. Note that if u and v are two elements of G, then $u \cdot g \cdot u^{-1}$ and $v \cdot g \cdot v^{-1}$ will represent the same element if, and only if,

$$\begin{aligned} u \cdot g \cdot u^{-1} = v \cdot g \cdot v^{-1} &\iff v^{-1} \cdot u \cdot g \cdot u^{-1} \cdot v = g \\ &\iff (v^{-1} \cdot u) \cdot g \cdot (v^{-1} \cdot u)^{-1} = g \\ &\iff v^{-1} \cdot u \in N_G(\{g\}) \\ &\iff u \in v \cdot N_G(\{g\}) \\ &\iff u \cdot N_G(\{g\}) = v \cdot N_G(\{g\}). \end{aligned}$$

Thus $u \cdot g \cdot u^{-1}$ and $v \cdot g \cdot v^{-1}$ represent the same element if, and only if, u and v belong to the same left coset of $N_G(\{g\})$. Therefore, to count all of the possible conjugates of g, we merely count the number of left cosets of $N_G(\{g\})$, which is

$$\frac{|G|}{|N_G(\{g\})|}.$$

□

We have already observed that the sum of the number of elements in each of the conjugacy classes must give the number of elements in the group. Since we now know how many elements are in each conjugacy class, we can derive what is called the *class equation.*

THEOREM 7.2: The Class Equation Theorem
Let G be a finite group. Then

$$|G| = \sum_{g} \frac{|G|}{|N_G(\{g\})|},$$

where the sum runs over one g from each conjugacy class.

PROOF We simply observe that every element of G appears in exactly one of the conjugacy classes. Thus, $|G|$ is the sum of the sizes of all of the conjugacy classes. We have by lemma 7.2 that the size of each conjugacy class is

$$\frac{|G|}{|N_G(\{g\})|}$$

where g is a representative element of the conjugacy class. Thus we get the class equation. □

We will see many very important applications of this equation, but let us begin by learning what this has to say about groups whose order is a power of a prime.

COROLLARY 7.2
If G is a group of order p^n where p is a prime and n is a positive integer, then $Z(G)$, the center of G, is not just the identity element.

PROOF First we observe that an element g is in the center of G if, and only if, $y \cdot g \cdot y^{-1} = g$ for all y in G, which would happen if, and only if, the conjugacy class of g consists of just g by itself.

Now suppose G is centerless. Then the only conjugacy class that contains just one element would be the class $\{e\}$. All other conjugacy classes would have a size that is a divisor of p^n, so the number of elements in the other conjugacy classes would be a power of p. But this is impossible since the sum on the right hand side of the class equation (7.2) would be congruent to 1 (Mod p), while the left hand side of the class equation would be p^n which is congruent to 0 (Mod p). Therefore, G is not centerless. □

This corollary is useful in finding all non-isomorphic groups of order p^n, where p is a prime. For example, we can easily find all non-isomorphic groups of order p^2.

COROLLARY 7.3

If p is a prime then there are exactly two non-isomorphic groups of order p^2, namely Z_{p^2} and $Z_p \times Z_p$.

PROOF If G is a group of order p^2, then by corollary 7.2, G has a nontrivial center. Since the number of elements of $Z(G)$ must divide p^2, so $|Z(G)|$ is either equal to p or p^2.

Suppose that $|Z(G)| = p$. Then there exists an element g not in $Z(G)$. Then $N_G(\{g\})$ denotes the set of elements that commute with g. Certainly

$$Z(G) \subseteq N_G(\{g\}),$$

and also

$$g \in N_G(\{g\}),$$

so $N_G(\{g\})$ contains at least $p+1$ elements. But this is a subgroup of G, so the number of elements must divide p^2. Hence, $N_G(\{g\})$ contains all of G, but this would say that g is in the center $Z(G)$, which contradicts our assumption. Thus, there are p^2 elements in $Z(G)$ and hence G is an abelian group.

Finally, we can use the fundamental theorem of finite abelian groups (6.2) to say that G must be isomorphic to the direct product of cyclic groups. It is easy to see that there are exactly two possibilities for such a product to have p^2 elements, namely Z_{p^2} and $Z_p \times Z_p$. □

In particular we can use corollary 7.3 to see that there are only two non-isomorphic groups of order 9, Z_9 and $Z_3 \times Z_3$.

One of the keys for finding all groups of a certain order is knowing whether there is a normal subgroup or a certain order. The next proposition will allow us to know that there will be a normal subgroup *without knowing the structure of the group.*

PROPOSITION 7.8

Let G be a group of order p^n. Then G contains a normal subgroup of order p^{n-1}.

PROOF We will proceed by using induction on n. Note that if $n = 1$, then there is obviously a normal subgroup of order $p^{1-1} = p^0 = 1$, namely the trivial subgroup $\{e\}$.

Suppose that we know that every group of order p^{n-1} has a normal subgroup of order p^{n-2}. Let G be a group of order p^n. Then by corollary 7.2, the center

of G is not just the identity element. Since p would then divide the order of $Z(G)$, by lemma 6.2 there is an element of $Z(G)$ of order p, say x. Then the group generated by x would be of order p, and since x is in the center, all elements of G would commute with x. Thus, $X = [x]$ would be a normal subgroup of G.

We then can consider the quotient group G/X. This would have order p^{n-1}, and we would have the canonical homomorphism

$$\phi : G \to G/X$$

whose kernel is the subgroup X. By the induction hypothesis, G/X is a group of order p^{n-1}, and so has a normal subgroup of order p^{n-2}, say Y.

We will now "lift" the subgroup Y back to the original group. Since $\phi^{-1}(Y)$ is the inverse image of a normal subgroup, by corollary 4.2, this is a normal subgroup of G. Note Y is a set of cosets, and that $g \in \phi^{-1}(Y)$ if, and only if, g is contained in one of the cosets of Y. Since each of the cosets of Y contains p elements, it is clear that the size of $\phi^{-1}(Y)$ is $p \cdot p^{n-2} = p^{n-1}$. Therefore, we have proved by induction that there is a normal subgroup of G of order p^{n-1}. □

We now are ready to start finding normal subgroups of a more general group, knowing only the group's order. The most important set of theorems that tackle this problem are by a Norwegian high school teacher named Ludwig Sylow (1832-1918). [1, p. 324] Before we work on finding normal subgroups let us see if we can find a subgroup of a given order within a group.

THEOREM 7.3: The First Sylow Theorem

Suppose that G is a group of order $p^n \cdot m$, where p is a prime, and m is coprime to p. Then G has a subgroup of order p^n.

PROOF We will proceed by using induction on the size of the group G. That is, we will assume that the theorem is true for all groups smaller than G.

If p^n divided $|H|$ for some proper subgroup H of G, then by our induction hypothesis, H would have a subgroup of order p^n, which would be a subgroup of G for which we are searching. So we may assume that p^n does not divide the order of any proper subgroup of G.

In particular, if g is not in the center of G, then $N_G(\{g\})$ will not be all of G. Hence, p^n does not divide $|N_G(\{g\})|$. But since p^n does divide $|G|$, we have from lemma 7.2 that the number of conjugates of g is $|G|/|N_G(\{g\})|$, which must be a multiple of p.

Now we can use the argument that we used in corollary 7.2. The class equation theorem (7.2) states that

$$|G| = \sum_g \frac{|G|}{|N_G(\{g\})|},$$

where the sum runs over one g from each conjugacy class. For those g in the center of G, $|G|/|N_G(\{g\})|$ will be 1, while for all other terms, $|G|/|N_G(\{g\})|$ will be a multiple of p. Since the sum is $p^n \cdot m$ which is a multiple of p, the number of elements in $Z(G)$ must be a multiple of p.

Since $Z(G)$ is an abelian group and p divides $Z(G)$, we have by lemma 6.2 that there is an element of $Z(G)$ of order p, say x. We now can proceed in the same way as we did in proposition 7.8. Since x is in the center, all elements of G would commute with x, and so $X = [x]$ would be a normal subgroup of order p.

The quotient group G/X would then have order $p^{n-1} \cdot m$, and we would have the canonical homomorphism

$$\phi : G \to G/X$$

whose kernel is the subgroup X. By the induction hypothesis, G/X is smaller than G, and so has a subgroup of order p^{n-1}, say Y. We can then lift Y back to the original group. Since $\phi^{-1}(Y)$ is the inverse image of a subgroup, by corollary 4.2, this is a subgroup of G. But the kernel of the homomorphism is of order p, so the size of $\phi^{-1}(Y)$ is $p \cdot p^{n-1} = p^n$. Therefore, we have proved by induction that there is a subgroup of G of order p^n. □

Since the first Sylow theorem guarantees the existance of at least one subgroup of order p^n for a group of size $p^n \cdot m$, we will give a name to these subgroups.

DEFINITION 7.5 If G is a group of order $p^n \cdot m$, where m is coprime to the prime p, then a subgroup of order p^n is called a *p-Sylow subgroup.*

Let us give a quick application of the first Sylow theorem (7.3). Suppose we have a group G of order 10. There is guaranteed to be a 2-Sylow subgroup, say H, and a 5-Sylow subgroup, say K. Obviously,

$$H \approx Z_2 \qquad \text{and} \qquad K \approx Z_5.$$

Furthermore, the intersection of H and K must just be the identity element, since Z_5 does not have any elements of order 2. Also, K is a subgroup of G with index 2, so by proposition 3.5, K is a normal subgroup of G. If H is also normal, we have by the direct product theorem (6.1) that

$$H \cdot K \approx H \times K \approx Z_2 \times Z_5 \approx Z_{10}.$$

On the other hand, if H is not a normal subgroup, then by the semi-direct product theorem (6.3)

$$H \cdot K \approx H \ltimes_{\phi} K$$

for some nontrivial homomorphism ϕ from H to $\mathrm{Aut}(K)$. But in chapter 6, we found that there was only one nontrivial homomorphism, yielding the dihedral group D_5. In either case, $H \cdot K$ is of order 10, so G is either isomorphic to Z_{10} or D_5.

Even though Sylow's first theorem (7.3) guarantees that there will be at least one p-Sylow subgroup, there may be more than one. The next of Sylow's theorems shows that any two p-Sylow subgroups are related.

THEOREM 7.4: The Second Sylow Theorem

If H and K are two p-Sylow subgroups of G, then there exists an element u in G such that $H = u \cdot K \cdot u^{-1}$.

PROOF Let G be a group of order $p^n \cdot m$, where m is coprime to the prime p. We begin by showing that whenever K is a p-Sylow subgroup of G then $u \cdot K \cdot u^{-1}$ will also be a p-Sylow subgroup for all u in G. Note that the number of elements in $u \cdot K \cdot u^{-1}$ is also p^n, and if $u \cdot k_1 \cdot u^{-1}$ and $u \cdot k_2 \cdot u^{-1}$ are two elements of $u \cdot K \cdot u^{-1}$, then

$$(u \cdot k_1 \cdot u^{-1}) \cdot (u \cdot k_2 \cdot u^{-1})^{-1} = u \cdot k_1 \cdot u^{-1} \cdot (u \cdot k_2^{-1} \cdot u^{-1}) = u \cdot (k_1 \cdot k_2^{-1}) \cdot u^{-1},$$

which is in $u \cdot K \cdot u^{-1}$. So by proposition 2.2, $u \cdot K \cdot u^{-1}$ is a p-Sylow subgroup of G.

If there is only one p-Sylow subgroup of G there is nothing to prove. Suppose H and K are two subgroups of order p^n. Let us call two elements u and v of G to be "related" if $u = h \cdot v \cdot k$ for some h in H and k in K. Note that every element is related to itself, for $u = e \cdot u \cdot e$, and e is in both H and K. Also, if u is related to v, then v is related to u, for

$$u = h \cdot v \cdot k \Longleftrightarrow v = h^{-1} \cdot u \cdot k^{-1}.$$

Finally, if u is related to v, and v is related to w, then $u = h_1 \cdot v \cdot k_1$ and $v = h_2 \cdot w \cdot k_2$, and so

$$u = h_1 \cdot (h_2 \cdot w \cdot k_2) \cdot k_1 = (h_1 \cdot h_2) \cdot w \cdot (k_2 \cdot k_1),$$

so u and w are related. Therefore, we can partition the group G into "families," where each family consists of all elements related to one element.

Now suppose that there are j families, and we select one element u_i from each family. Each of the families can be described as $H \cdot u_i \cdot K$. Hence, we can write

$$G = (H \cdot u_1 \cdot K) \cup (H \cdot u_2 \cdot K) \cup \cdots \cup (H \cdot u_j \cdot K).$$

Since each of the families have no elements in common, we have

$$|G| = |H \cdot u_1 \cdot K| + |H \cdot u_2 \cdot K| + \cdots + |H \cdot u_j \cdot K|.$$

How many elements are in each family? We note that $H \cdot u_i \cdot K$ has the same number of elements as $H \cdot u_i \cdot K \cdot u_i^{-1}$. We saw that $u_1 \cdot K \cdot u_i^{-1}$ is a group, and so even though the product of two groups was not always a group, proposition 4.9 gave us the number of elements in the set to be

$$|H \cdot u_i \cdot K| = |H \cdot u_i \cdot K \cdot u_i^{-1}| = \frac{|H| \cdot |u_i \cdot K \cdot u_i^{-1}|}{|H \cap (u_i \cdot K \cdot u_i^{-1})|} = \frac{p^n \cdot p^n}{|H \cap (u_i \cdot K \cdot u_i^{-1})|}.$$

If we plug this formula into the equation above it, we have that

$$p^n \cdot m = \frac{p^n \cdot p^n}{|H \cap (u_1 \cdot K \cdot u_1^{-1})|} + \frac{p^n \cdot p^n}{|H \cap (u_2 \cdot K \cdot u_2^{-1})|} + \cdots + \frac{p^n \cdot p^n}{|H \cap (u_j \cdot K \cdot u_j^{-1})|}.$$

Note that the intersection of two groups is a subgroup of both the groups, and so the denominators will all be powers of p. Dividing both sides of the equation by p^n, we have

$$m = \frac{p^n}{|H \cap (u_1 \cdot K \cdot u_1^{-1})|} + \frac{p^n}{|H \cap (u_2 \cdot K \cdot u_2^{-1})|} + \cdots + \frac{p^n}{|H \cap (u_j \cdot K \cdot u_j^{-1})|}.$$

Since m is not a multiple of p, there must be some term on the right hand side of this equation that is not a multiple of p. But this can happen only if one of the denominators is p^n, that is,

$$|H \cap (u_i \cdot K \cdot u_i^{-1})| = |H|$$

for some i. Since H and $u_i \cdot K \cdot u_i^{-1}$ both have p^n elements, we must have $H = u_i \cdot K \cdot u_i^{-1}$. Therefore, for any two p-Sylow subgroups of G, there is a u such that $H = u \cdot K \cdot u^{-1}$. □

The second Sylow theorem (7.4) allows us to know exactly when a p-Sylow subgroup is normal.

COROLLARY 7.4

The group G has only one p-Sylow subgroup for a given prime p if, and only if, G has a p-Sylow subgroup that is normal.

PROOF Suppose that H is the only p-Sylow subgroup of G. Then for any element u in G, $u \cdot H \cdot u^{-1}$ will be a p-Sylow subgroup of G. But since there is only one p-Sylow subgroup, we have $u \cdot H \cdot u^{-1} = H$ for all u in G. Hence, H is a normal subgroup.

Now suppose that H is a normal p-Sylow subgroup of G. By the second Sylow theorem (7.4) every other p-Sylow subgroup is of the form $u \cdot H \cdot u^{-1}$. But since H is normal, $u \cdot H \cdot u^{-1} = H$. Therefore, H is the only p-Sylow subgroup. □

The natural question that corollary 7.4 raises is, "How do we know if there is only one p-Sylow subgroup?" The next lemma allows us to find the number of p-Sylow subgroups in terms of the size of the normalizer. In fact it allows us to find the number of p-Sylow subgroups of a certain type.

LEMMA 7.3
Let G be a group of order $p^n \cdot m$, and let P be a p-Sylow subgroup of G. Let H be any other subgroup of G. Then the number of p-Sylow subgroups that can be written as $u \cdot P \cdot u^{-1}$ with u an element of H is given by

$$\frac{|H|}{|N_G(P) \cap H|}.$$

PROOF Since P is a subgroup of G, $N_G(P)$ is a subgroup of G, so the intersection of $N_G(P)$ and H will be a subgroup of G. We can use the same argument as lemma 7.2, and note that if u and v are two elements of H, then $u \cdot P \cdot u^{-1}$ and $v \cdot P \cdot v^{-1}$ will represent the same p-Sylow subgroup if, and only if,

$$\begin{aligned} u \cdot P \cdot u^{-1} = v \cdot P \cdot v^{-1} &\iff v^{-1} \cdot u \cdot P \cdot u^{-1} \cdot v = P \\ &\iff (v^{-1} \cdot u) \cdot P \cdot (v^{-1} \cdot u)^{-1} = P \\ &\iff v^{-1} \cdot u \in N_G(P) \cap H \\ &\iff u \in v \cdot (N_G(P) \cap H) \\ &\iff u \cdot (N_G(P) \cap H) = v \cdot (N_G(P) \cap H). \end{aligned}$$

Thus, $u \cdot P \cdot u^{-1}$ and $v \cdot P \cdot v^{-1}$ represent the same p-Sylow subgroup if, and only if, $u \cdot (N_G(P) \cap H)$ and $v \cdot (N_G(P) \cap H)$ are the same left cosets of $N_G(P) \cap H$. Therefore, the number of p-Sylow subgroups that can be expressed as $u \cdot P \cdot u^{-1}$, with u an element of H, is

$$\frac{|H|}{|N_G(P) \cap H|}.$$ □

We now are ready to prove the last of Sylow's theorem, which in many cases will tell us the number of p-Sylow subgroups of a group.

THEOREM 7.5: The Third Sylow Theorem
Suppose that the number of p-Sylow subgroups of G is k. Then k divides $|G|$, and $k \equiv 1 \pmod{p}$.

PROOF Suppose that we label the p-Sylow subgroups of G as P_0, P_1, $P_2, \ldots, P_{k-1}$. Let us partition all of the p-Sylow subgroups of G into different categories where two p-Sylow subgroups P_i and P_j are in the same category if there is an element u in P_0 such that

$$P_j = u \cdot P_i \cdot u^{-1}.$$

Note that P_0 would be in its own category while the number of p-Sylow subgroups in the other categories would be, according to lemma 7.3,

$$\frac{|P_0|}{|N_G(P_i) \cap P_0|}$$

where P_i is one p-Sylow subgroup in the category.

Recall that the normalizer of each P_i contains P_i as a normal subgroup, so $N_G(P_i)$ is divisible by p^n, and hence by corollary 7.4 the only p-Sylow subgroup of $N_G(P_i)$ is P_i. Thus, the intersection of $N_G(P_i)$ with P_0 is smaller than P_0 when $i > 0$. Since the order of P_0 is p^n, we have that the number of p-Sylow subgroups in each category, besides the category containing just P_0, is a power of p, and hence is a multiple of p.

Therefore, the total number of p-Sylow subgroups is one more than a multiple of p, so $k \equiv 1 \pmod{p}$.

Finally, if we let $H = G$ in lemma 7.3, we find that the number of conjugates of P_0 is

$$\frac{|G|}{|N_G(P_0)|}.$$

By the second Sylow theorem (7.4), this would give us all of the p-Sylow subgroups. Therefore, k is also a divisor of the order of the group G. □

These three theorems of Sylow provide a means of finding normal subgroups of a group G just from knowing the order of G. For example, suppose that a group is of order 45. Since 3^2 divides 45, there is a 3-Sylow subgroup of order 9. We also know that the number of 3-Sylow subgroups divides 45, so this number must be 1, 3, 5, 9, 15, or 45. However, the number must be congruent to 1 (Mod 3). Thus, the only possibility is that there is only one subgroup of order 9, say H. But then this subgroup is normal.

We can use the same argument to find a normal subgroup of order 5. Again, the number of 5-Sylow subgroups must be 1, 3, 5, 9, 15, or 45. But this number must also be congruent to 1 (Mod 5), so there is only one subgroup of order 5, and this group must also be normal.

Although the Sylow theorems are powerful tools, when combined with the tools of semi-direct products and the computational power of GAP or *Mathematica*, we can determine most of the groups of a given order. For example, let us see if we can find all of the groups of order 12.

If G is a group of order 12, since the divisors of 12 are 1, 2, 3, 4, 6, and 12, by the third Sylow theorem there are either one or four 3-Sylow subgroups and

there are either one or three 2-Sylow subgroups. Let H be a 3-Sylow subgroup, and let K be a 2-Sylow subgroup (which will be of order 4). Certainly the intersection of H and K is just the identity element since K cannot contain an element of order 3.

Let us show that either H or K is normal. If H is not normal, there must be four 3-Sylow subgroups of G. Each of these 3-Sylow groups contains two different elements of order 3, so G would have eight elements of order 3. But that would leave only four elements left over, and so K must be composed of all of those four elements. Then there would be only one 2-Sylow subgroup, which would be normal.

By the direct product theorem (6.1) and the semi-direct product theorem (6.3), $H \cdot K$ would have to be of one of the following forms:

1. $H \cdot K \approx Z_3 \times Z_4 \approx Z_{12}$,
2. $H \cdot K \approx Z_3 \times Z_8^* \approx Z_3 \times Z_2 \times Z_2$,
3. $H \cdot K \approx Z_3 \ltimes_\phi Z_4$,
4. $H \cdot K \approx Z_3 \ltimes_\phi Z_8^*$,
5. $H \cdot K \approx Z_4 \ltimes_\phi Z_3$,
6. $H \cdot K \approx Z_8^* \ltimes_\phi Z_3$.

In all six cases $H \cdot K$ contains 12 elements, and so $G = H \cdot K$. Let us work these six cases separately. The first two give the two possible abelian groups of order 12. Case 3 is actually impossible, since $\mathrm{Aut}(Z_4) \approx Z_4^*$ has only two elements, and therefore has no elements of order 3. Therefore, there is no nontrivial homomorphism from Z_3 to $\mathrm{Aut}(Z_4)$. The other three cases are as follows:

Case 4

An element of order 3 in Z_3 must map to an element of order 3 in $\mathrm{Aut}(Z_8^*)$, which is isomorphic to S_3. There are two elements of order 3 in S_3, and these two elements are conjugates. By proposition 6.7, it does not matter which element of Z_3 maps to which elements in $\mathrm{Aut}(Z_8^*)$, so the semi-direct product $Z_3 \ltimes_\phi Z_8^*$ is unique up to isomorphisms. But A_4 is a group of order 12, has a normal subgroup isomorphic to Z_8^*, and does not have a normal subgroup of order 3. Thus, A_4 must be this unique semi-direct product $Z_3 \ltimes Z_8^*$.

Case 5

The homomorphism ϕ must map a generator of Z_4 to a nontrivial element of $\mathrm{Aut}(Z_3)$. But $\mathrm{Aut}(Z_3)$ has only two elements, so this homomorphism is uniquely determined. The group is generated by the *Mathematica* commands

InitGroup[e];
Define[a^3, e]; Define[b^4, e]
Define[1/a, a^2]; Define[1/b, b^3]
Define[b.a, a.a.b]
M = Group[{a, b}]

or the GAP commands

```
gap> f := FreeGroup("a","b");;
gap> a := f.1;; b:=f.2;;
gap> g := f/[a^3,b^4,b*a/(a*a*b)];;
gap> NumberElements := true;
true
gap> MultTable(g);

*       |1  2  3  4  5  6  7  8  9  10 11 12
--------+------------------------------------
e       |1  2  3  4  5  6  7  8  9  10 11 12
b       |2  3  5  6  1  7  9  10 4  11 12 8
b^2     |3  5  1  7  2  9  4  11 6  12 8  10
a       |4  10 7  8  12 2  11 1  5  6  3  9
b^3     |5  1  2  9  3  4  6  12 7  8  10 11
a^2*b   |6  11 9  10 8  3  12 2  1  7  5  4
a*b^2   |7  12 4  11 10 5  8  3  2  9  1  6
a^2     |8  6  11 1  9  10 3  4  12 2  7  5
a^2*b^3 |9  8  6  12 11 1  10 5  3  4  2  7
a*b     |10 7  12 2  4  11 5  6  8  3  9  1
a^2*b^2 |11 9  8  3  6  12 1  7  10 5  4  2
a*b^3   |12 4  10 5  7  8  2  9  11 1  6  3
```

From the multiplication table, this non-abelian group has only one element of order 2. Thus, it is not isomorphic to any group we have seen before. If we ask GAP for the description of the structure,

```
gap> StructureDescription(g);
"C3 : C4"
```

which can be interpreted as $Z_4 \ltimes Z_3$. This is how we will identify this group.

Case 6

Since $\mathrm{Aut}(Z_3)$ contains only two elements, the homomorphism ϕ is completely determined by its kernel. The kernel of ϕ cannot be just the identity, since there is not an isomorphic copy of Z_8^* in $\mathrm{Aut}(Z_3)$. On the other hand, the kernel of a nontrivial homomorphism cannot be all of Z_8^*. Thus, the kernel contains exactly two elements, and because there are automorphisms of Z_8^* mapping one subgroup of order 2 to any other, it will not matter which subgroup of order 2 we pick. Thus, there is a unique semi-direct product $Z_8^* \ltimes Z_3$.

The obvious group of order 12 that we have yet to consider is $Z_2 \times S_3$. This has a normal subgroup of order 3, so by process of elimination must be $Z_8^* \ltimes Z_3$. In summary, we have found five possible groups of order 12:

$$Z_{12}, \qquad A_4 \qquad Z_2 \times Z_2 \times Z_3 \qquad Z_2 \times S_3 \quad \text{and} \quad Z_4 \ltimes Z_3.$$

Let us summarize our findings formally with a proposition.

PROPOSITION 7.9
There are exactly 28 non-isomorphic groups of order less than 16.

PROOF The trivial group is the only group of order 1, and since 2, 3, 5, 7, 11, and 13 are prime, we have only one non-isomorphic group of each of these orders.

In chapter 4 we found that the only non-isomorphic groups of order 4 were

$$Z_4 \quad \text{and} \quad Z_8^*,$$

the only non-isomorphic groups of order 6 were

$$Z_6 \quad \text{and} \quad S_3,$$

and the only non-isomorphic groups of order 8 were

$$Z_8, \quad Z_{15}^*, \quad Z_{24}^*, \quad Q, \quad \text{and} \quad D_4.$$

By corollary 7.3 the only two non-isomorphic groups of order 9 are

$$Z_9 \quad \text{and} \quad Z_3 \times Z_3.$$

We have already used the first Sylow theorem (7.3) to find all of the non-isomorphic groups of order 10:

$$Z_{10} \quad \text{and} \quad D_5.$$

We just found all of the groups of order 12:

$$Z_{12}, \quad A_4, \quad Z_2 \times Z_2 \times Z_3, \quad Z_2 \times S_3, \quad \text{and} \quad Z_4 \ltimes Z_3.$$

We can use the same argument to find all of the non-isomorphic groups of order 14. If $|G| = 14$, there must be a 7-Sylow subgroup of G, say K. Since K contains half the elements, by proposition 3.5, K is normal. We also must have a 2-Sylow subgroup, H. Since K cannot have an element of order 2, H and K have only the identity element in common. If H is normal, then $H \cdot K \approx H \times K \approx Z_2 \times Z_7 \approx Z_{14}$. If H is not normal, by the semi-direct product theorem (6.3),

$$H \cdot K \approx H \ltimes_{\phi} K$$

for some homomorphism ϕ from H to $\mathrm{Aut}(K)$. In either case $H \cdot K$ has 14 elements, and so $G = H \cdot K$. Also, ϕ is determined by where the non-identity element of H is mapped. Since this must be an element of $\mathrm{Aut}(K)$ of order 2, and since

$$\mathrm{Aut}(K) \approx \mathrm{Aut}(Z_7) \approx Z_7^* \approx Z_6$$

has only one element of order 2, there can only be one such homomorphism. Since D_7 is a non-abelian group of order 14, this must be the one semi-direct product that we found. Thus, the only two groups of order 14 are

$$Z_{14} \quad \text{and} \quad D_7.$$

Let us move on to find all groups of order 15. Suppose $|G| = 15$. Then the number of 3-Sylow subgroups and the number of 5-Sylow subgroups must both divide 15, so both of these numbers must be one of 1, 3, 5, or 15. But 1 is the only number in this set that is congruent to 1 (Mod 5). So there is only one 5-Sylow subgroup, K. Likewise, 1 is the only number in the set that is congruent to 1 (Mod 3). So there is only one 3-Sylow subgroup, H. By corollary 7.4, both K and H are normal subgroups of G, and the intersection must be just the identity element. Thus, by corollary 6.1,

$$H \cdot K \approx H \times K \approx Z_3 \times Z_5 \approx Z_{15}.$$

Since this has all 15 elements, this must be all of G, and so there is only one non-isomorphic group of order 15, namely Z_{15}.

Therefore, counting all of the groups of order less than 16, we find that there are exactly 28 of them. □

Unfortunately, finding all the groups of order 16 is a difficult problem. Even though proposition 7.8 tells us that there must be a normal subgroup K of order 8, there is no guarantee that there would be a subgroup H of order 2 such that $H \cdot K$ gives the whole group. Thus, we would not be able to use the semi-direct product theorem (6.3) to find *all* of the groups of order 16 (although we can find many of them, as we did in the last chapter).

Problems for Chapter 7

Interactive Problems

7.1 Use *Mathematica* or GAP to find the center of the group D_6. This can be loaded in *Mathematica* by

```
InitGroup[e];
Define[a^2, e]; Define[b^6, e]
```

Define[b.a, a.b.b.b.b.b]
Define[1/a, a]; Define[1/b, b^5]
D6 = Group[{a, b}]

or in GAP by

```
gap> f := FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> D6 := f/[a^2,b^6,b*a/(a*b^5)];; a := D6.1;; b := D6.2;;
```

What familiar group is the quotient group $D_6/Z(D_6)$ isomorphic to?

7.2 Use *Mathematica* or GAP to find the normalizer $N_{D_6}(\{x\})$ for each of the 12 elements of the group D_6 listed in problem 7.1. For which elements is the normalizer the same subgroup?

7.3 Use *Mathematica*'s or GAP's `NormalClosure` command to find all of the normal subgroups of the group D_6 given in problem 7.1.

7.4 The following commands load a group of order 20 into *Mathematica*.

InitGroup[e];
Define[a^4, e]; Define[b^5, e]
Define[1/a, a^3]; Define[1/b, b^4]
Define[b.a, a.b.b]
M = Group[{a, b}]

Here are the commands to load the same group in GAP:

```
gap> f := FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> M := f/[a^4,b^5,b*a/(a*b^2)];; a := M.1;; b := M.2;;
```

Find the conjugacy classes of this group, and use this to find all of the normal subgroups of M.

7.5 Use *Mathematica* or GAP to find all of the 2-Sylow and 5-Sylow subgroups of the group M defined in problem 7.4. How many of the subgroups are there? Does this agree with the prediction given by the third Sylow theorem?

7.6 Using GAP or *Mathematica*, find all non-isomorphic groups of order 21.
Hint: What can you determine from Sylow's theorems? Which semi-direct products are possible?

Non-Interactive Problems

7.7 Find the center of the group D_4.

7.8 Find the center of the group D_5.

7.9 For each element g in D_4, find the normalizer $N_{D_4}(\{g\})$.

7.10 For each element g in D_5, find the normalizer $N_{D_5}(\{g\})$.

7.11 Must the center of a group be abelian?

7.12 Must the normalizer of an element $N_G(\{g\})$ be abelian?

7.13 Find all of the conjugacy classes of the group D_4.

7.14 Find all of the conjugacy classes of the quaternion group Q. (See table 4.3 in chapter 4 for the multiplication table of Q.)

7.15 Find all of the conjugacy classes of the group D_5.

7.16 Let G be a group and $Z(G)$ the center of G. Prove that G is abelian if, and only if, $G/Z(G)$ is cyclic.

Hint: Use proposition 7.3.

7.17 Let G be any group. Prove that

$$Z(G) = \bigcap_{g \in G} N_G(\{g\}).$$

7.18 Let G be a group, and let g be an element of G. Prove that

$$N_G(\{g\}) = N_G(\{g^{-1}\}).$$

7.19 *Mathematica* and GAP showed that the group A_5 had conjugacy classes of orders 1, 12, 12, 15, and 20. Using this information alone, without using Abel's theorem (7.1), prove that A_5 is simple.

Hint: A normal subgroup must contain the union of several conjugacy classes, including $\{e\}$. But the number of elements must satisfy Lagrange's theorem (3.1).

7.20 GAP showed that the group $\mathrm{Aut}(Z_{24}^*)$ had conjugacy classes of orders 1, 21, 24, 24, 42, and 56. Using this information alone, prove that $\mathrm{Aut}(Z_{24}^*)$ is simple.

7.21 The group A_6 has seven conjugacy classes of orders 1, 40, 40, 45, 72, 72, and 90. With this information alone, without using Abel's theorem (7.1), prove that A_6 is simple.

7.22 The group $L_2(8)$ has 504 elements, and has nine conjugacy classes of orders 1, 56, 56, 56, 56, 63, 72, 72, and 72. Prove that $L_2(8)$ is simple. This is another example of a Chevalley group.

7.23 The group $L_2(11)$ has 660 elements, and has eight conjugacy classes of orders 1, 55, 60, 60, 110, 110, 132, and 132. Prove this group is simple. This group, the fifth smallest non-cyclic simple group, is related to the group $\mathrm{Aut}(Z_{11} \times Z_{11})$.

7.24 The group M_{11} has order 7920, and has 10 conjugacy classes of orders 1, 165, 440, 720, 720, 990, 990, 990, 1320, and 1584. Prove that M_{11} is simple. This is the smallest of the 26 sporadic simple groups.

7.25 The group $L_3(4)$ has 20160 elements, and has 10 conjugacy classes of orders 1, 315, 1260, 1260, 1260, 2240, 2880, 2880, 4032, and 4032. Prove that this group is simple. Show that even though A_8 is a simple group with the same order, these two groups are not isomorphic.

Hint: How many 3-cycles are in A_8? What does lemma 7.1 say about the 3-cycles?

7.26 Find a representative element for each of the seven conjugacy classes of the group A_6. The number of elements in each conjugacy class is given in problem 7.21.

Hint: Are (12345) and (12354) in the same conjugacy class? Why are (12)(3456) and (12)(3465) in the same conjugacy class?

7.27 Using the counting methods used to estimate the 168 elements of $\mathrm{Aut}(Z_{24}^*)$, find the maximum number of elements of $\mathrm{Aut}(Z_2 \times Z_2 \times Z_2 \times Z_2)$. This group is in fact simple, and contains the number of elements predicted by this estimate. Are there any other simple groups that we have seen of this order?

7.28 If G has order p^n for some prime p, show that every subgroup of order p^{n-1} is a normal subgroup of G.

7.29 If H is a subgroup of G, and H has order p^i for some prime p, show that H is contained in a p-Sylow subgroup of G.

Hint: Mimic the proof of the second Sylow theorem (7.4).

7.30 Use Sylow's theorem to show that all groups of order 33 are cyclic.

7.31 Prove that no group of order 56 is simple.

7.32 Show that if p is an odd prime, then any group with $2p$ elements is isomorphic to either Z_{2p} or D_p.

7.33 Determine all non-isomorphic groups of order 99.

7.34 Show that there are exactly four non-isomorphic groups of order 66:

$$Z_{66}, \qquad D_{33}, \qquad D_{11} \times Z_3, \quad \text{and} \quad D_3 \times Z_{11}.$$

Hint: Use Sylow's theorems along with problem 7.30.

7.35 Show that all groups of order 255 are cyclic.
Hint: Use lemma 4.5.

7.36 Let $|G| = p \cdot q$, where $p > q$ are both primes. Show that G has a normal subgroup of order p.

7.37 If $|G| = p^2 \cdot q$, where p and q are different primes, show that G must contain a normal subgroup of either size p^2 or q.
Hint: Generalize the case $|G| = 12$ done in the text.

7.38 Show that a group of order $p^3 \cdot q$, where p and q are different primes, cannot be simple.
Hint: Use corollary 5.2 for the case $|G| = 24$. Then do the case $q < p$. With these out of the way, you can assume that $q > p + 1$.

7.39 Use the results of problems 7.36 through 7.38 to show that no non-cyclic group of order less than 60 is simple.

This page intentionally left blank

Chapter 8

Solvable and Insoluble Groups

8.1 Subnormal Series and the Jordan-Hölder Theorem

In this chapter we will study the concept of *solvable* groups. But first we must make some preliminary definitions. We have already encountered situations in which we had a normal subgroup of a normal subgroup, such as in the second isomorphism theorem. But suppose we have a whole series of subgroups of a group G, each one fitting inside of the previous one like Russian dolls.

DEFINITION 8.1 A *subnormal series* for a group G is a sequence $G_0, G_1, G_2, \ldots G_n$ of subgroups of G such that

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{e\},$$

where each G_i is a normal subgroup of G_{i-1} for $i = 1, 2, \cdots n$.

A subnormal series is called a *normal series* if it satisfies the stronger condition that all of the groups G_i are normal subgroups of the original group G. We will be mainly interested in subnormal series, but there are a few of the exercises regarding normal series.

The group S_4, for example, has a normal subgroup of order 4, namely

K = Group[{P[2,1,4,3], P[4,3,2,1]}]

```
gap> K := Group( (1,2)(3,4), (1,4)(2,3) );
Group([ (1,2)(3,4) ])
gap> List(K);
[ (), (1,2)(3,4), (1,4)(2,3), (1,3)(2,4) ]
```

The identity element is of course a normal subgroup of K, so we can write

$$S_4 \supseteq K \supseteq \{(\,)\}$$

which would be a subnormal series of length $n = 2$. Is there a way that we can make a longer series out of this one? Because A_4 is also a normal subgroup of S_4, and K is a normal subgroup of A_4, we can slip this group into our series. Also, the group K contains the subgroup

H = Group[{ P[2,1,4,3]}]

```
gap> H := Group( (1,2)(3,4));
Group([ (1,2)(3,4) ])
gap> List(H);
[ (), (1,2)(3,4) ]
```

which is a normal subgroups of K since K is abelian. Therefore, we have a longer subnormal series of length 4:

$$S_4 \supseteq A_4 \supseteq K \supseteq H \supseteq \{()\}.$$

We say that this new subnormal series is a *refinement* of the first subnormal series.

DEFINITION 8.2 We say that a subnormal (or normal) series

$$G = H_0 \supseteq H_1 \supseteq H_2 \supseteq \cdots \supseteq H_k = \{e\}$$

is a *refinement* of the subnormal (or normal) series

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{e\}$$

if each subgroup G_i appears as H_j for some j.

Is there a way that we can refine our subnormal series to produce an even longer chain? Our definition did not exclude the possibility of two groups in the series being the same, so we could consider

$$S_4 \supseteq A_4 \supseteq A_4 \supseteq K \supseteq H \supseteq H \supseteq H \supseteq \{P[\ ]\}.$$

Although this is a longer subnormal series, it is usually pointless to repeat the same subgroup in the series.

DEFINITION 8.3 A *composition series* of a group G is a subnormal series

$$G = G_0 \supset G_1 \supset G_2 \supset \cdots \supset G_n = \{e\}$$

for which each subgroup is smaller than the proceeding subgroup, and for which there is no refinement that includes additional subgroups.

There is a GAP command `CompositionSeries` that finds one possible composition series for a given group.

```
gap> S4 := Group( (1,2), (2,3,4) );
Group([ (1,2), (2,3,4) ]);
gap> L := CompositionSeries(S4);
[ Group([ (3,4), (2,4,3), (1,3)(2,4), (1,2)(3,4) ]),
```

```
  Group([ (2,4,3), (1,3)(2,4), (1,2)(3,4) ]),
  Group([ (1,3)(2,4), (1,2)(3,4) ]), Group([ (1,2)(3,4) ]),
  Group(()) ]
gap> List(L, Size);
[ 24, 12, 4, 2, 1 ]
```

GAP selected the composition series

$$S_4 \supseteq A_4 \supseteq K \supseteq H \supseteq \{()\}.$$

We see that since no subgroups are repeated, and there simply is not enough room between two of these subgroups to slip in another subgroup, that this indeed is a composition series for S_4. In fact, we can easily test to see whether a subnormal series is a composition series.

PROPOSITION 8.1
The subnormal series

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{e\}$$

is a composition series if, and only if, all of the quotient groups G_{k-1}/G_k are nontrivial simple groups.

PROOF Note that if there are no repeated subgroups in the subnormal series then G_{i-1}/G_i must contain at least two elements. Likewise, if G_{i-1}/G_i is nontrivial, then G_{i-1} is not equal to G_i. So the quotient groups are nontrivial if, and only if, there are no repeated subgroups in the subnormal series.

Suppose that the subnormal series is not a composition series yet does not repeat any subgroups. Then there must be an additional group H that we can add between G_{k-1} and G_k, so that

$$G_{k-1} \supseteq H \supseteq G_k,$$

where H is a normal subgroup of G_{k-1} and G_k is a normal subgroup of H. Then by lemma 4.3, H/G_k will be a normal subgroup of G_{k-1}/G_k, and since H is neither G_{k-1} nor G_k, we have a proper normal subgroup of G_{k-1}/G_k.

Now suppose that there is a proper normal subgroup N of G_{k-1}/G_k. Can we then lift N to find a suitable subgroup H to fit between G_{k-1} and G_k? If we consider the canonical homomorphism ϕ from G_{k-1} to the quotient group G_{k-1}/G_k we can take $H = \phi^{-1}(N)$. Then since N is a normal subgroup of G_{k-1}/G_k, by corollary 4.2 H will be a normal subgroup of G_{k-1}. Also, G_k will be a normal subgroup of H, for H is in G_{k-1}. Because N has at least two elements, H will be strictly larger than the kernel of ϕ, yet since N is not the entire image of ϕ, H will be strictly smaller than G_k. Therefore, the subnormal series is not a composition series.

Thus, a subnormal series is a composition series if, and only if, the quotient groups G_{k-1}/G_k are nontrivial simple groups. □

The quotient groups G_{k-1}/G_k in a composition series for G are called the *composition factors* of the composition series.

For example, the composition factors for the composition series

$$S_4 \supseteq A_4 \supseteq K \supseteq H \supseteq \{(\,)\}$$

are

$$S_4/A_4 \approx Z_2, \qquad A_4/K \approx Z_3, \qquad K/H \approx Z_2, \quad \text{and} \quad H/\{(\,)\} \approx Z_2.$$

These are displayed in GAP by the command `DisplayCompositionSeries`.

```
gap> DisplayCompositionSeries(S4);
G (4 gens, size 24)
 | Z(2)
S (3 gens, size 12)
 | Z(3)
S (2 gens, size 4)
 | Z(2)
S (1 gens, size 2)
 | Z(2)
1 (0 gens, size 1)
```

It is certainly possible for a group to have more than one composition series. For example, we could have picked the subgroup $B = \{(\,), (1,4)(2,3)\}$, given in *Mathematica*® by

B = Group[{ P[4,3,2,1]}]

instead of H, producing the composition series

$$S_4 \supseteq A_4 \supseteq K \supseteq B \supseteq \{(\,)\}.$$

Even though this is a different composition series, the composition factors are isomorphically the same. Our goal for this section is to prove that this happens all of the time. However, we have yet to see why two composition series must have the same length. Even if we can prove that the composition series are the same length, the composition factors may not appear in the same order. For example, the group Z_{12} has the following two subnormal series:

$$Z_{12} \supseteq \{0,3,6,9\} \supseteq \{0\}.$$

$$Z_{12} \supseteq \{0,2,4,6,8,10\} \supseteq \{0,4,8\} \supseteq \{0\}.$$

No matter how we refine these series, the quotient group isomorphic to Z_3 in the first series will come before any other nontrivial quotient groups, yet any refinement of the second series will have the last nontrivial quotient group isomorphic to Z_3.

It helps if we use a diagram to demonstrate the strategy that we will be using. Suppose that we have a group G with two subnormal series, one of length 2, and one of length 3, as pictured in figure 8.1.

$$G = A_0 \supseteq A_1 \supseteq A_2 = \{e\}, \qquad G = B_0 \supseteq B_1 \supseteq B_2 \supseteq B_3 = \{e\}.$$

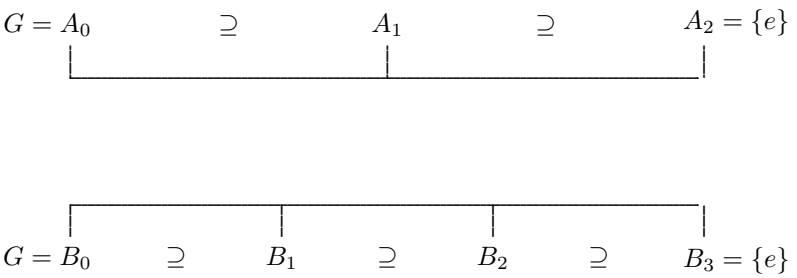


FIGURE 8.1: Two subnormal series of different lengths

It is immediately clear that $A_0 = B_0$ and $A_2 = B_3$, but A_1 does not have to be either B_1 or B_2.

The goal is to refine both of the subnormal series by adding two subgroups within each gap of the A series, and one subgroup within each gap in the B series. Here, we will allow the possibility of duplicate subgroups in the refinements. Nonetheless, both series will have length 6, which we can express as follows:

$$G = A_0 \supseteq A_{1,1} \supseteq A_{1,2} \supseteq A_1 \supseteq A_{2,1} \supseteq A_{2,2} \supseteq A_0 = \{e\},$$

$$G = B_0 \supseteq B_{1,1} \supseteq B_1 \supseteq B_{1,2} \supseteq B_2 \supseteq B_{1,3} \supseteq B_0 = \{e\}.$$

Figure 8.2 shows these set inclusions, and also gives a hint on how we are to define these intermediate subgroups.

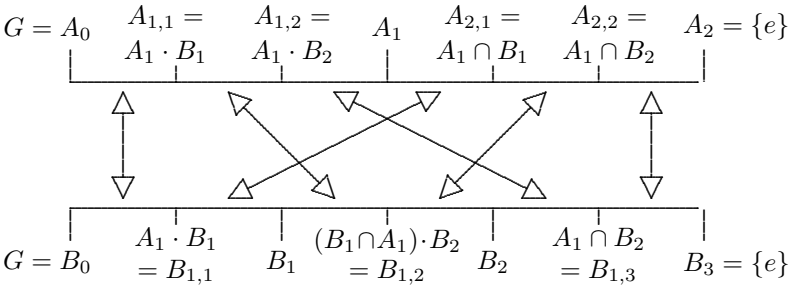


FIGURE 8.2: Strategy for the refinement theorem

The next step will be to show that the quotient groups for each interval of the A series is isomorphic to a quotient group for an interval of the B series, as shown by the arrows in figure 8.2. Note that this scrambles the order of the quotient groups, so that the i-th subinterval of the j-th interval in the A series corresponds to the j-th subinterval of the i-th interval of the B series.

Although it is clear that

$$G \supseteq A_1 \cdot B_1 \supseteq A_1 \cdot B_2 \supseteq A_1 \supseteq A_1 \cap B_1 \supseteq A_1 \cap B_2 \supseteq \{e\}, \qquad \text{and}$$
$$G \supseteq A_1 \cdot B_1 \supseteq B_1 \supseteq (B_1 \cap A_1) \cdot B_2 \supseteq B_2 \supseteq A_1 \cap B_2 \supseteq \{e\},$$

it is not at all clear that each is a normal subgroup of the previous group, or even that all of these sets are subgroups of G. Before we show this, we will need the following lemma.

LEMMA 8.1
Let X, Y, and Z be three subgroups of the group G, with Y being a subgroup of X, and $Y \cdot Z = Z \cdot Y$. Then

$$X \cap (Y \cdot Z) = Y \cdot (X \cap Z) = (X \cap Z) \cdot Y.$$

PROOF Note that $(X \cap Z) \subseteq X$, and since $Y \subseteq X$, $Y \cdot (X \cap Z) \subseteq X$. Also, $(X \cap Z) \subseteq Z$, so $Y \cdot (X \cap Z) \subseteq Y \cdot Z$. Hence,

$$Y \cdot (X \cap Z) \subseteq X \cap (Y \cdot Z).$$

All we need to do is prove the inclusion in the other direction. Suppose that $x \in X \cap (Y \cdot Z)$. Then x is in X, and can also be written as $x = y \cdot z$, where y is in Y, and z is in Z. But then $z = y^{-1} \cdot x$ would be in both X and Z. Thus,

$$x = y \cdot (y^{-1} \cdot x) \in Y \cdot (X \cap Z).$$

Therefore, we have inclusions in both directions, so

$$Y \cdot (X \cap Z) = X \cap (Y \cdot Z).$$

So far, we haven't used the fact that $Y \cdot Z = Z \cdot Y$. By lemma 4.4, $Y \cdot Z$ is a subgroup of G, and so the intersection of X with $Y \cdot Z$ is a subgroup of G. So by lemma 4.4 again, we have

$$Y \cdot (X \cap Z) = (X \cap Z) \cdot Y. \qquad \square$$

We will need one more lemma that will help us to show the isomorphisms indicated by the arrows in figure 8.2.

LEMMA 8.2
Let X, Y, and Z be three subgroups of the group G, with Y being a normal subgroup of X, and Z a normal subgroup of G. Then $Y \cdot Z$ is a normal subgroup of $X \cdot Z$, and

$$(X \cdot Z)/(Y \cdot Z) \approx X/(X \cap (Y \cdot Z)).$$

PROOF Since Z is a normal subgroup of G, both $Y \cdot Z$ and $X \cdot Z$ are subgroups of G by lemma 4.5. If we let $y \cdot z$ be in $Y \cdot Z$, and $x \cdot w$ be in $X \cdot Z$, then

$$\begin{aligned}(x \cdot w) \cdot (y \cdot z) \cdot (x \cdot w)^{-1} &= x \cdot (y \cdot x^{-1} \cdot x \cdot y^{-1}) \cdot w \cdot y \cdot z \cdot w^{-1} \cdot x^{-1} \\ &= (x \cdot y \cdot x^{-1}) \cdot (x \cdot (y^{-1} \cdot w \cdot y) \cdot z \cdot w^{-1} \cdot x^{-1}).\end{aligned}$$

Now, $x \cdot y \cdot x^{-1}$ is in Y, since Y is a normal subgroup of X. Likewise, $y^{-1} \cdot w \cdot y$ is in Z, since y is in G. Then $(y^{-1} \cdot w \cdot y) \cdot z \cdot w^{-1}$ is in Z, and so $x \cdot (y^{-1} \cdot w \cdot y) \cdot z \cdot w^{-1} \cdot x^{-1}$ is in Z, since x is in G. Therefore, $(x \cdot w) \cdot (y \cdot z) \cdot (x \cdot w)^{-1}$ is in $Y \cdot Z$, and so $Y \cdot Z$ is a normal subgroup of $X \cdot Z$.

We now can use the third isomorphism theorem (4.3), using $K = Y \cdot Z$. We have that $X \cdot K = X \cdot Y \cdot Z = X \cdot Z$ since Y is a subgroup of X. So

$$(X \cdot Z)/(Y \cdot Z) = (X \cdot K)/K \approx X/(X \cap K) = X/(X \cap (Y \cdot Z)). \qquad \square$$

We are now ready to put the pieces together, and show any two subnormal series can be refined in such a way that the quotient groups are isomorphic.

THEOREM 8.1: The Refinement Theorem

Suppose that there are two subnormal series for a group G. That is, there are subgroups A_i and B_j such that

$$G = A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n = \{e\},$$

and

$$G = B_0 \supseteq B_1 \supseteq B_2 \supseteq \cdots \supseteq B_m = \{e\},$$

where each A_i is a normal subgroup of A_{i-1}, and each B_j is a normal subgroup of B_{j-1}. Then it is possible to refine both series by inserting the subgroups

$$A_{i-1} = A_{i,0} \supseteq A_{i,1} \supseteq A_{i,2} \supseteq \cdots \supseteq A_{i,m} = A_i, \qquad i = 1, 2, \ldots n,$$

$$B_{j-1} = B_{j,0} \supseteq B_{j,1} \supseteq B_{j,2} \supseteq \cdots \supseteq B_{j,n} = B_j, \qquad j = 1, 2, \ldots m$$

in such a way that

$$A_{i,j-1}/A_{i,j} \approx B_{j,i-1}/B_{j,i}.$$

PROOF We let

$$A_{i,j} = (A_{i-1} \cap B_j) \cdot A_i \quad \text{and} \quad B_{j,i} = (B_{j-1} \cap A_i) \cdot B_j.$$

To see that these fit the conditions we need, we first want to show that these are groups. Note that both

$$X = (A_{i-1} \cap B_{j-1}) \quad \text{and} \quad Y = (A_{i-1} \cap B_j)$$

are subgroups of A_{i-1}, Y is a subgroup of X, and $Z = A_i$ is a normal subgroup of A_{i-1}.

So by lemma 4.5, both $A_{i,j-1} = X \cdot Z$ and $A_{i,j} = Y \cdot Z$ are subgroups of A_{i-1}. We can now use lemma 8.2, using $G = A_{i-1}$. Since B_j is a normal subgroup of B_{j-1}, Y is a normal subgroup of X, so by lemma 8.2, $Y \cdot Z$ is a normal subgroup of $X \cdot Z$, and

$$A_{i,j-1}/A_{i,j} = (X \cdot Z)/(Y \cdot Z) \approx X/(X \cap (Y \cdot Z)).$$

Now lemma 8.1 comes into use. Since Y is a subgroup of X,

$$\begin{aligned} X \cap (Y \cdot Z) = Y \cdot (X \cap Z) &= (A_{i-1} \cap B_j) \cdot (A_{i-1} \cap B_j \cap A_i) \\ &= (A_{i-1} \cap B_j) \cdot (A_i \cap B_{j-1}) \\ &= (A_i \cap B_{j-1}) \cdot (A_{i-1} \cap B_j). \end{aligned}$$

Thus,

$$A_{i,j-1}/A_{i,j} \approx (A_{i-1} \cap B_{j-1})/[(A_{i-1} \cap B_j) \cdot (A_i \cap B_{j-1})].$$

By switching the roles of the two series we find by the exact same argument that

$$B_{j,i-1}/B_{j,i} \approx (B_{j-1} \cap A_{i-1})/[B_{j-1} \cap A_i) \cdot (B_j \cap A_{i-1})].$$

Notice that these are exactly the same thing, so

$$A_{i,j-1}/A_{i,j} \approx B_{j,i-1}/B_{j,i}. \qquad \square$$

If we now apply the refinement theorem to two composition series we find that the composition factors will be the same.

THEOREM 8.2: The Jordan-Hölder Theorem
Let G be a finite group, and let

$$G = A_0 \supset A_1 \supset A_2 \supset \cdots \supset A_n = \{e\}$$

and

$$G = B_0 \supset B_1 \supset B_2 \supset \cdots \supset B_m = \{e\}$$

be two composition series for G. Then $n = m$, and the composition factors A_{u-1}/A_u are isomorphic to the composition factors B_{v-1}/B_v in some order.

PROOF By the refinement theorem (8.1), there is a refinement of both composition series such that the quotient groups of the two subnormal series are isomorphic to each other in some order. In particular, the nontrivial quotient groups of one subnormal series are isomorphic to the nontrivial quotient groups of the other. But these are composition series, so any refinements

merely repeat a subgroup a number of times. Thus, by eliminating these repetitions, we eliminate the trivial quotient groups and produce the original two composition series. Thus, the quotient groups A_{u-1}/A_u are isomorphic to the quotient groups B_{v-1}/B_v in some order. The fact that $n = m$ merely comes from the one-to-one correspondence of the nontrivial quotient groups. □

The Jordan-Hölder theorem (8.2) shows that the composition factors do not depend on the composition series, but rather the finite group G. This is reminiscent of the unique factorization of integers, where every integer greater than one can be written as a unique product of prime numbers. Since the composition factors are always nontrivial simple groups, in a sense the simple groups play the same role in group theory that prime numbers play in number theory. The correspondence is heightened by the fact that Z_p is a nontrivial simple group if, and only if, p is a prime number. However, we have seen that there are other simple groups, such as $\mathrm{Aut}(Z_{24}^*)$ and A_n for $n > 4$. Since these groups are rather large (at least 60 elements), they will only show up as composition factors for very large groups.

For example, a composition series for S_5 is given by

$$S_5 \supset A_5 \supset \{()\}, \qquad S_5/A_5 \approx Z_2, \quad \text{and} \quad A_5/\{()\} \approx A_5.$$

```
gap> S5 := Group( (1,2),(2,3,4,5) );
Group([ (1,2), (2,3,4,5) ]);
gap> CompositionSeries(S5);
[ Group([ (1,2), (2,3,4,5) ]),
  Group([ (1,3,2), (1,4,3), (1,4,5) ]), Group(()) ]
```

Since Z_2 and A_5 are both simple groups, this is a composition series, and so the composition factors of S_5 are Z_2 and A_5.

The composition series will play a vital role in determining whether groups are solvable or not. However, we will hold off on the definition of a solvable group until we have defined another tool in group theory, the derived group.

8.2 Derived Group Series

In this section we will find a method for producing a composition series that is easily implemented using *Mathematica* or GAP.

DEFINITION 8.4 Given two elements x and y of a group G, the *commutator* of x and y is the element $x^{-1} \cdot y^{-1} \cdot x \cdot y$, and is written $[x, y]$.

Notice that if G is an abelian group the commutator will always give the identity element. We can also consider the commutator of two subgroups of

G. If H and K are two subgroups, then consider the set

$$\{x^{-1} \cdot y^{-1} \cdot x \cdot y \quad | \quad x \in H \quad \text{and} \quad y \in K\}.$$

Unfortunately, this set will not always form a group. The simplest example is found in S_4. We can consider the two subgroups

$$H = \{(\,), (1\,2)\}, \qquad K = \{(\,), (2\,3\,4), (2\,4\,3)\}.$$

Then the set

$$\{x^{-1} \cdot y^{-1} \cdot x \cdot y \quad | \quad x \in H \quad \text{and} \quad y \in K\}$$

can be found by making a table for possible values of x and y.

$x^{-1} \cdot y^{-1} \cdot x \cdot y$	$(\,)$	$(2\,3\,4)$	$(2\,4\,3)$
$(\,)$	$(\,)$	$(\,)$	$(\,)$
$(1\,2)$	$(\,)$	$(1\,2\,3)$	$(1\,2\,4)$

So we get $\{(\,), (1\,2\,3), (1\,2\,4)\}$, which is not a subgroup. However, we can consider the group *generated* by all of the commutators, which of course will make a subgroup.

DEFINITION 8.5 Given two subgroups H and K of a group G, we define the *mutual commutator subgroup of H and K,* denoted $[H, K]$, to be the subgroup generated by the elements

$$\{x^{-1} \cdot y^{-1} \cdot x \cdot y \quad | \quad x \in H \quad \text{and} \quad y \in K\}.$$

We can find the mutual commutator with the *Mathematica* commands

```
H = Group[{C[1,2]}];
K = Group[{C[2,3,4]}];
MutualCommutator[H, K]
```

or the GAP commands

```
gap> H := Group((1,2));;
gap> K := Group((2,3,4));;
gap> C := CommutatorSubgroup(H,K);
Group([ (1,2,3), (1,4,3) ])
gap> Size(C);
12
```

So the commutator $[H, K]$ in this case is A_4. Note that whenever an element u is in $[H, K]$, we cannot say that $u = x^{-1} \cdot y^{-1} \cdot x \cdot y$ for some $x \in H$ and $y \in K$. Rather, we must write

$$u = u_1 \cdot u_2 \cdot \cdots \cdot u_n,$$

where either u_i or u_i^{-1} is $x_i^{-1} \cdot y_1^{-1} \cdot x_i \cdot y_i$. In spite of this difficulty, we will be able to discover some important properties with the mutual commutator groups.

PROPOSITION 8.2
If H and K are normal subgroups of G, then $[H, K]$ is a normal subgroup of G.

PROOF Let u be an element of $[H, K]$, and v an element of G. Then $u = u_1 \cdot u_2 \cdots u_n$, where either u_i or $u_i - 1$ is $x_i^{-1} \cdot y_i^{-1} \cdot x_i \cdot y_i$. Then

$$v \cdot u \cdot v^{-1} = (v \cdot u_1 \cdot v^{-1}) \cdot (v \cdot u_2 \cdot v^{-1}) \cdots (v \cdot u_n \cdot v^{-1}),$$

and

$$\begin{aligned} v \cdot x_i^{-1} \cdot y_i^{-1} \cdot x_i \cdot y_i \cdot v^{-1} = \\ (v \cdot x_i^{-1} \cdot v^{-1}) \cdot (v \cdot y_i^{-1} \cdot v^{-1}) \cdot (v \cdot x_i \cdot v^{-1}) \cdot (v \cdot y_i \cdot v^{-1}) = \\ \left[v \cdot x_i \cdot v^{-1}, v \cdot y_i \cdot v^{-1}\right]. \end{aligned}$$

If H and K are both normal subgroups of G, then $v \cdot x_i \cdot v^{-1}$ is in H, and $v \cdot y_i \cdot v^{-1}$ is in K. Thus, $[v \cdot x_i \cdot v^{-1}, v \cdot y_i \cdot v^{-1}]$ is in $[H, K]$. Since $(v \cdot u_i \cdot v^{-1})^{-1} = (v \cdot u_i^{-1} \cdot v^{-1})$, if one of these is in $[H, K]$, they both are. Hence $v \cdot u_i \cdot v^{-1}$ is in $[H, K]$ for every u_i, and $v \cdot u \cdot v^{-1} \in [H, K]$. By proposition 3.4, $[H, K]$ is a normal subgroup of G. □

Many times one of the two groups H or K will be the whole group G. We call the subgroup $[G, H]$ the *commutator subgroup of H in G*. In this case *Mathematica* can find the commutator subgroup faster with the simplified command

Commutator[G, H]

which takes advantage of the fact that H is a subgroup of G. In fact, *Mathematica* will correctly find the commutator subgroup if only the *generators* of H are specified. For example, suppose we wish to find the commutator $[S_4, A_4]$.

```
S4 = Group[{ C[1,2], C[1,2,3,4] }]
A4 = Group[{ C[1,2,3], C[2,3,4] }]
```

It is faster to use only the generators of A_4:

```
Commutator[S4,{ C[1,2,3], C[2,3,4] }]
```

which gives us A_4 again. The commutator $[S_4, S_4]$ is given by

```
Commutator[S4,{ C[1,2], C[1,2,3,4] }]
```

which is also A_4. However, the commutator $[A_4, A_4]$ is

Commutator[A4,{ C[1,2,3], C[2,3,4] }]

which gives a subgroup with only four elements. This is exactly the subgroup K from the last section. The GAP commands for this are

```
gap> S4 := Group( (1,2), (2,3,4) );
gap> A4 := Group( (1,2,3), (2,3,4) );
gap> List(CommutatorSubgroup(S4,A4) );
[ (), (1,2,3), (1,3,2), (1,4)(2,3), (2,3,4), (1,2)(3,4),
  (1,3,4), (1,4,3), (2,4,3), (1,2,4), (1,3)(2,4), (1,4,2) ]
gap> List(CommutatorSubgroup(S4,S4) );
[ (), (1,2,3), (1,3,2), (1,4)(2,3), (2,3,4), (1,2)(3,4),
  (1,3,4), (1,4,3), (2,4,3), (1,2,4), (1,3)(2,4), (1,4,2) ]
gap> List(CommutatorSubgroup(A4,A4) );
[ (), (1,4)(2,3), (1,2)(3,4), (1,3)(2,4) ]
```

DEFINITION 8.6 We define the commutator subgroup of G with itself, $[G, G]$, to be the *derived group* of G, denoted G'.

Since G is a normal subgroup of itself, proposition 8.2 states that the derived group will be a normal subgroup of G. Since the commutator of any two elements in an abelian group is e, $[G, G]$ will be the trivial group whenever G is abelian.

We can denote the derived group of the derived group G' as G''. Likewise, the derived group of G'' will be denoted G''', and so on. Because each of these groups is a normal subgroup of the previous one, we have the series

$$G \supseteq G' \supseteq G'' \supseteq G''' \supseteq \cdots.$$

This is called the *derived series* for the group G. The derived series is in fact a subnormal series as long as the groups keep getting smaller and smaller until they finally get to the trivial subgroup. In GAP, we can use the shorter command `DerivedSubgroup` for $[G, G]$. For example, the derived group series of $G = S_4$ is

```
gap> Gp := DerivedSubgroup(S4);
Group([ (1,3,2), (1,4,3) ])
gap> List(Gp);
[ (), (1,2,3), (1,3,2), (1,4)(2,3), (2,3,4), (1,2)(3,4),
  (1,3,4), (1,4,3), (2,4,3), (1,2,4), (1,3)(2,4), (1,4,2) ]
gap> Gpp := DerivedSubgroup(Gp);
Group([ (1,4)(2,3), (1,2)(3,4) ])
gap> List(Gpp);
[ (), (1,4)(2,3), (1,2)(3,4), (1,3)(2,4) ]
gap> Gppp := DerivedSubgroup(Gpp);
Group(())
gap> List(Gppp);
[ () ]
```

So $G' = A_4$, $G'' = K$, and $G''' = \{(\,)\}$, since K is abelian. So we produce the series

$$S_4 \supseteq A_4 \supseteq K \supseteq \{(\,)\}.$$

However, if we start with the group A_5, then $[A_5, A_5]$ must be a normal subgroup of the simple group A_5. Since the derived group is not the identity element, we see that the derived group is all of A_5.

```
gap> A5 := Group( (1,2,3),(3,4,5) );
Group([ (1,2,3), (3,4,5) ])
gap> Size(DerivedSubgroup(A5) );
60
```

Thus, the derived series for A_5 is

$$A_5 \supseteq A_5 \supseteq A_5 \supseteq A_5 \supseteq \cdots$$

which never gets to the trivial subgroup.

DEFINITION 8.7 A group G is called *solvable* if the derived series

$$G \supseteq G' \supseteq G'' \supseteq G''' \supseteq \cdots$$

includes the trivial group in a finite number of steps. If the derived series never reaches the trivial group, G is said to be *insoluble.*

By our experiments, we see that S_4 is a solvable group, whereas A_5 is not. In fact that GAP command `IsSolvable` bears this out.

```
gap> IsSolvable(S4);
true
gap> IsSolvable(A5);
false
```

Whenever we have a solvable group G, the derived series is in fact a subnormal series for G. So it is natural that the derived series would shed some light as to what the composition factors of G are. First we will need the following lemma, which characterizes the derived group.

LEMMA 8.3
Let G be a group. Then the derived group G' is the smallest normal subgroup for which the quotient group is abelian.

PROOF First we need to show that G/G' is abelian. Consider the canonical homomorphism ϕ from G onto G/G'. Then for x and y in G, $x^{-1} \cdot y^{-1} \cdot x \cdot y$ is in G', and so $\phi(x^{-1} \cdot y^{-1} \cdot x \cdot y)$ is the identity element in G/G'. But then

$$\phi(x^{-1} \cdot y^{-1} \cdot x \cdot y) = \phi(x)^{-1} \cdot \phi(y)^{-1} \cdot \phi(x) \cdot \phi(y) = e,$$

so $\phi(x) \cdot \phi(y) = \phi(y) \cdot \phi(x)$. Since ϕ is surjective, we see that G/G' is abelian.

Now suppose that N is another normal subgroup of G for which G/N is abelian. To show that G' is a smaller group, we will show that N contains G'.

For any x and y in G, note that $x^{-1} \cdot y^{-1} \cdot x \cdot y$ is certainly contained in $x^{-1} \cdot N \cdot y^{-1} \cdot N \cdot x \cdot N \cdot y \cdot N$. But since the quotient group G/N is abelian, we have

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

Thus, $x^{-1} \cdot y^{-1} \cdot x \cdot y$ is in N for all x and y in G. Since G' is generated by all such elements, G' is contained in N. □

We now can express a relationship between the composition factors of a group and the derived series of a group.

THEOREM 8.3: The Solvability Theorem

Let G be a finite group. Then G is solvable if, and only if, the composition factors of G are cyclic groups of prime order.

PROOF Suppose that the composition factors of G are all cyclic groups of prime order. Then there exists a composition series for G:

$$G = G_0 \supset G_1 \supset G_2 \supset \cdots \supset G_n = \{e\}.$$

Since G_0/G_1 is an abelian group, we have from lemma 8.3 that G' is contained in G_1. But since G_1/G_2 is also abelian, by lemma 8.2 we have G_1' is in G_2, and so

$$G'' \subseteq G_1' \subseteq G_2.$$

Proceeding in this way we find that the n-th derived group, $G^{(n)}$, must be contained in $G_n = \{e\}$. Thus, the derived series produced the trivial group in at most n steps, so G is solvable.

Now suppose that G is solvable and finite, and so the derived series can be written

$$G \supseteq G' \supseteq G'' \supseteq G''' \supseteq \cdots \supseteq G^{(n)} = \{e\}.$$

If $G^{(n)}$ is the first term in the derived series equal to $\{e\}$, then this subnormal series can never repeat any two subgroups. Because this is a finite group, there are only a finite number of ways this series could be refined without repeating subgroups. Thus, by the refinement theorem, we can refine this to produce a composition series. Because each of the quotient groups of the derived series is abelian, the quotient groups of the refinement must also be abelian. But by proposition 8.1, the quotient groups of the composition series must be nontrivial simple groups. The only nontrivial simple groups that are abelian are the cyclic groups of prime order. Thus, the quotient groups for

this composition series are cyclic groups of prime order. By the Jordan-Hölder theorem (8.2), all composition series are the same way. □

From the solvability theorem we see that for finite groups, solvability can be defined in terms of the composition factors. Does this hold true for infinite groups as well? That is, is an infinite group solvable as long as there is no non-abelian simple group (finite or infinite) lurking somewhere within its structure, either as a subgroup or as a quotient group? To shed some light on this problem, we will first need the following lemma.

LEMMA 8.4
If N is a normal subgroup of G, and H is a subgroup of G, then

$$(H \cdot N/N)' = (H' \cdot N)/N.$$

PROOF We first note that since N is a normal subgroup of G, $H \cdot N$ is a subgroup of G, and so N is a normal subgroup of $H \cdot N$. Two typical elements of $H \cdot N/N$ are $h \cdot n \cdot N$ and $k \cdot m \cdot N$, where h and k are in H, and n and m are in N. Then $(H \cdot N/N)'$ is generated from the elements of the form

$$(h \cdot n \cdot N)^{-1} \cdot (k \cdot m \cdot N)^{-1} \cdot (h \cdot n \cdot N) \cdot (k \cdot m \cdot N) = h^{-1} \cdot k^{-1} \cdot h \cdot k \cdot N.$$

But these elements are also in $(H' \cdot N)/N$. In fact, $(H' \cdot N)/N$ is generated by the elements of the form $h^{-1} \cdot k^{-1} \cdot h \cdot k \cdot N$. Therefore, the groups $(H \cdot N/N)'$ and $(H' \cdot N)/N$ are equal. □

With this lemma we will be able to show the relationship with a solvable group to its subgroups and quotient groups.

PROPOSITION 8.3
Suppose that G is a group and H is a normal subgroup of G. Then G is solvable if, and only if, both H and G/H are solvable.

PROOF We begin by showing that if G is solvable, and H is a subgroup of G, normal or not, then H is solvable. Since H is contained in G, we have

$$H' \subseteq G' \Longrightarrow H'' \subseteq G'' \Longrightarrow H''' \subseteq G''' \cdots.$$

Thus, since $G^{(n)} = \{e\}$ for some n, $H^{(n)} = \{e\}$, and H is solvable.

Next we want to show that if H is normal, then G/H is solvable. Since $G = G \cdot H$ we can use lemma 8.4 to find $(G/H)' = (G' \cdot H)/H$. But since G' is a subgroup, we can continue to use lemma 8.4 to find

$$(G/H)'' = (G' \cdot H/H)' = (G'' \cdot H)/H,$$

$$(G/H)''' = (G'' \cdot H/H)' = (G''' \cdot H)/H, \cdots.$$

Since G is a solvable group, $G^{(n)} = \{e\}$ for some n. Thus

$$(G/H)^{(n)} = (G^{(n)} \cdot H)/H$$

would be the identity group H/H. Therefore, G/H is a solvable group.

Now suppose that both H and G/H are solvable. Then $(G/H)^n$ is the identity for some n, so $(G^{(n)} \cdot H)/H$ is the identity. Thus, $G^{(n)}$ is a subgroup of H, and since H is solvable, $G^{(n)}$ must be solvable. Therefore, $G^{(n+m)}$ is the identity for some m, and so G is a solvable group. □

From this proposition, we see that for an infinite solvable group there cannot be any non-abelian simple groups within its structure whether as a subgroup, a quotient group, a subgroup of a quotient group, etc. Thus the current definition of solvability for infinite groups agrees with the historical notion of a group that does not contain non-abelian simple groups in the composition factors.

Why do we want to know whether a group is solvable or not? Notice that the *solvable* groups could be entered into *Mathematica* using the **InitGroup** and **Define** commands, whereas the *insoluble* groups, such as $\mathrm{Aut}(Z_{24}^*)$, had to be considered as a subgroup of a symmetric group. In the next section, we will show why the solvable groups were the only groups that could be entered into *Mathematica* using the **Define** commands.

8.3 Polycyclic Groups

Throughout these notebooks, we used *Mathematica*'s **InitGroup** and **Define** commands or GAP's `FreeGroup` command to produce many of the groups we have been studying. Only occasionally did we have to use permutations to represent groups, such as the groups A_5 and $\mathrm{Aut}(Z_{24}^*)$. However, the method for converting a finite group into a set of *Mathematica* or GAP commands has never been fully explained. We know that the groups can be represented by a small number of generators. Why was S_4 defined in *Mathematica* with three generators when only two generators would generate the group?

The method for defining a group G in *Mathematica* using a set of generators stems from the composition series for a solvable group G. However, a composition series is actually more than we need. We will still insist that the factors of a series be cyclic, but not necessarily of prime order.

DEFINITION 8.8 A subnormal series

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{e\}$$

is a *polycyclic series* if the quotient groups G_{i-1}/G_i are all cyclic groups. The number n is called the *length* of the polycyclic series.

It is obvious that a group with a polycyclic series must be solvable, since the cyclic quotient groups would be solvable. Although any finite solvable group has a polycyclic series, it should be noted that an *infinite* solvable group may not always have a polycyclic series. The groups that have a polycyclic series are called *polycyclic groups.*

Given a polycyclic series for a polycyclic group,

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{e\},$$

we can find a set of generators and relationships between the generators that will allow us to define the group in *Mathematica* or GAP. Since G_{i-1}/G_i is cyclic, we can choose an element $g_i \in G_{i-1}$ such that $g_i G_i$ is a generator of G_{i-1}/G_i. Then if G_{i-1}/G_i has order n_i, then $g_i^{n_i} \in G_i$. Also, if $j > i$, then $[g_j, g_i] \in [G_{i-1}, G_{i-1}] \subseteq G_i$. Since $j > i$, $g_j \in G_i$, so we have that $g_j \cdot g_i \in g_i G_i$. This means that for each pair $1 \leq i < j \leq n$, we can define a relation of the form

$$g_j \cdot g_i = g_i \cdot (\text{ element of } G_i).$$

This definition would allow *Mathematica* or GAP to unravel a combination of generators that are "in the wrong order." That is, if we consider the generators $g_1, g_2, \ldots g_n$ as "letters," going in alphabetical order, then these definitions would find a way of expressing the element of the group as a product of generators such that the generators are in alphabetical order.

In *Mathematica*, the groups must be defined in a form similar to a polycyclic representation. In fact, the groups defined using **InitGroup** and **Define** so far are either polycyclic representations, or a mirror image of such a representation. For example, if we wish to use a polycyclic series to define the group Q, we could use

$$G_0 = Q \supseteq G_1 = \{1, i, -1, -i\} \supseteq G_2 = \{1\},$$

and let $g_1 = j$ and $g_2 = i$. Since G_0/G_1 is of order 2, we know that g_1^2 is in G_1, and indeed $j^2 = -1 = i^2$. Also, $i^4 = 1 \in G_2$. Finally, we need to compute $[g_2, g_1] = [i, j] = -1 = i^2$. Thus, $i^{-1} \cdot j^{-1} \cdot i \cdot j = i^2$, so $i \cdot j = j \cdot i^3$. Thus, the commands

```
InitGroup[e];
Define[i^4, e]
Define[j^2 , i^2]
Define[i.j, j.i.i.i]
Q = Group[j, i]
```

will define the group Q. This puts the elements in "alphabetical" order, because $g_1 = j$ is considered to be before $g_2 = i$. Of course, it makes more

sense to have i come before j, so we can take the "mirror image" of this definition

InitGroup[e];
Define[i^4, e]
Define[j^2 , i^2]
Define[j.i, i.i.i.j]
Q = Group[i, j]

which of course will define an isomorphically equivalent group.

Here is a more complicated example. We have a polycyclic series for S_4,

$$G_0 = S_4 \supseteq G_1 = A_4 \supseteq G_2 = K \supseteq G_3 = H \supseteq G_4 = \{(\,)\}$$

and we would like to enter this into *Mathematica* or GAP using generators. Since there are four cyclic quotient groups, we will need four generators g_1, g_2, g_3, g_4 such that $g_i G_i$ is a generator of G_{i-1}/G_i. Some obvious choices are $g_1 = (1,2)$, $g_2 = (1,2,3)$, $g_3 = (1,3)(2,4)$, and $g_4 = (1,2)(3,4)$.

Next, $g_i^{n_i} \in G_i$, where n_i is the order of G_{i-1}/G_i. Looking at the polycyclic series for S_4, we find that $n_1 = 2$, $n_2 = 3$, $n_3 = 2$, and $n_4 = 2$. Hence we calculate $g_1^2 = (\,)$, $g_2^3 = (\,)$, $g_3^2 = (\,)$, and $g_4^2 = (\,)$. In this case, all of these turned out to be the identity element, but we are only promised that $g_i^{n_i}$ will be in G_i, and hence expressible in terms of $g_{i+1}, \ldots g_n$.

Finally, we calculate $[g_j, g_i] \in G_i$ for each combination $j > i$, and express each of these in terms of $g_{i+1}, \ldots g_n$. We find that $[g_2, g_1] = (1\,2\,3) = g_2$, $[g_3, g_1] = (1\,2)(3\,4) = g_4$, $[g_4, g_1] = (\,)$, $[g_3, g_2] = (1\,4)(2\,3) = g_3 \cdot g_4$. $[g_4, g_2] = (1\,3)(2\,4) = g_3$, and $[g_4, g_3] = (\,)$.

We are now ready to enter this into GAP as a polycyclic group. We can use a, b, c, and d as the four generators, and use GAP's `Comm` command for the commutator of two elements.

```
gap> f:= FreeGroup("a","b","c","d");;
gap> a:= f.1;; b:=f.2;; c:=f.3;; d:=f.4;;
gap> g:=f/[a^2,b^3,c^2,d^2, Comm(b,a)/b, Comm(c,a)/d, Comm(d,a),
> Comm(c,b)/(c*d), Comm(d,b)/c, Comm(d,c) ];
<fp group on the generators [ a, b, c, d ]>
gap> List(g);
[ <identity ...>, a, b, c, a*b*a*c*b, a*b, a*c, b*a*c*b, a*b*a,
  b*c, c*b, a*c*a, b*a, a*b*c, a*c*b, c*a, a*b*a*c, b*c*b,
  b*a*c*a, b*a*c, a*b*c*b, a*b*a*c*a, a*b*c*a, b*c*a ]
```

GAP is expressing each element as a product of generators, but not always in alphabetical order. But since we used a polycyclic series to define this group, we can convert it to a polycyclic form with the `PcGroupFpGroup` command. This converts an fp group (defined using commutators as we did) to pc groups.

```
gap> h := PcGroupFpGroup(g);
<pc group of size 24 with 4 generators>
gap> a := h.1;; b:= h.2;; c:= h.3;; d:=h.4;;
```

```
gap> List(h);
[ <identity> of ..., d, c, c*d, b, b*d, b*c, b*c*d, b^2, b^2*d,
  b^2*c, b^2*c*d, a, a*d, a*c, a*c*d, a*b, a*b*d, a*b*c,
  a*b*c*d, a*b^2, a*b^2*d, a*b^2*c, a*b^2*c*d ]
```

Now every element besides the identity is expressed as a product of generators in alphabetical order. GAP can work with polycyclic groups (pc groups) much more efficiently than with general groups defined using the `FreeGroup` command. In fact, very often GAP will express a group as a polycyclic group by default.

Here is another example. Table 8.1 shows a multiplication table for a non-abelian group that we will simply call A.

TABLE 8.1: Multiplication table for the mystery group A

·	1	Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L
1	1	Z	Y	X	W	V	U	T	S	R	Q	P	O	N	M	L
Z	Z	Y	X	1	T	W	V	U	R	Q	P	S	L	O	N	M
Y	Y	X	1	Z	U	T	W	V	Q	P	S	R	M	L	O	N
X	X	1	Z	Y	V	U	T	W	P	S	R	Q	N	M	L	O
W	W	V	U	T	S	R	Q	P	O	N	M	L	1	Z	Y	X
V	V	U	T	W	P	S	R	Q	N	M	L	O	X	1	Z	Y
U	U	T	W	V	Q	P	S	R	M	L	O	N	Y	X	1	Z
T	T	W	V	U	R	Q	P	S	L	O	N	M	Z	Y	X	1
S	S	R	Q	P	O	N	M	L	1	Z	Y	X	W	V	U	T
R	R	Q	P	S	L	O	N	M	Z	Y	X	1	T	W	V	U
Q	Q	P	S	R	M	L	O	N	Y	X	1	Z	U	T	W	V
P	P	S	R	Q	N	M	L	O	X	1	Z	Y	V	U	T	W
O	O	N	M	L	1	Z	Y	X	W	V	U	T	S	R	Q	P
N	N	M	L	O	X	1	Z	Y	V	U	T	W	P	S	R	Q
M	M	L	O	N	Y	X	1	Z	U	T	W	V	Q	P	S	R
L	L	O	N	M	Z	Y	X	1	T	W	V	U	R	Q	P	S

Because there are no elements of order 8, this cannot be one of the groups of the form $Z_2 \ltimes_\phi Z_8$ studied in section 6.4.

Finding a polycyclic series is not hard, but finding a short series of length 2 is a little trickier. We find that $\{1, Z, Y, X\}$ is a normal subgroup isomorphic to Z_4, and the quotient group is also cyclic. Thus, the series

$$G_0 = A \supset G_1 = \{1, Z, Y, X\} \supset G_2 = \{1\}$$

is a polycyclic series of length 2. By using this series, we need only two generators, a and b. Since G_1/G_2 has two generators, $\{Z\}$ and $\{X\}$, we can

let b represent either element, say $b = Z$. Then $b^4 = Z^4$ must be in $G_2 = \{1\}$, so

InitGroup[e];
Define[b^4, e]

defines $b = Z$ in *Mathematica*. Next, we notice that both $\{W, V, U, T\}$ and $\{O, N, M, L\}$ are generators of G_0/G_1. Thus, we can let a be any of these eight elements, say $a = W$. Then $a^4 = W^4$ must be in G_1, and in fact the table shows that $a^4 = e$.

Define[a^4, e]

Finally, we need to let *Mathematica* know how to handle the combination $b \cdot a$. We know that the commutator $[b, a]$ is in G_1, and using the multiplication table we have that $b^{-1} \cdot a^{-1} \cdot b \cdot a = Z^{-1} \cdot W^{-1} \cdot Z \cdot W = Y = b^2$. So $b \cdot a = a \cdot b^3$. While we are at it, we can also define the inverses of the two generators a and b.

Define[b.a, a.b.b.b]
Define[1/a, a^3]
Define[1/b, b^3]
A = Group[{a, b}]

This same strategy can be used to define this group as a pc group in GAP.

```
gap> f:= FreeGroup("a","b");; a := f.1;; b:=f.2;;
gap> g:= f/[a^4, b^4, Comm(a,b)/b^2];;
gap> h:= PcGroupFpGroup(g);
#I  You are creating a Pc group with non-prime relative orders.
#I  Many algorithms require prime relative orders.
#I  Use 'RefinedPcGroup' to convert.
<pc group of size 16 with 2 generators>
gap> a := h.1;; b:= h.2
gap> List(h);
[ <identity> of ..., b, b^2, b^3, a, a*b, a*b^2, a*b^3, a^2,
  a^2*b, a^2*b^2, a^2*b^3, a^3, a^3*b, a^3*b^2, a^3*b^3 ]
```

GAP gives a warning that we did not use a composition series to define the group, and so some of the features will not be available to us. Of course, using a composition series would require four generators, and hence more work. Most of the operations will still work for this group, such as multiplication tables, but to analyze the group

```
gap> StructureDescription(g);
"C4 : C4"
```

we have to use the fp version. We see that this group is a semi-direct product of Z_4 with itself. In fact, it is the only such semi-direct product, so we can refer to this group as $Z_4 \ltimes Z_4$.

Both GAP's pc groups and *Mathematica*'s groups are rewriting systems. That is, the fundamental methodology is to replace certain combinations of generators with other combinations until no more possible replacements are possible. But there is still one question that has not been addressed. How do we know for certain that the computer will not get hung in a loop? Consider the following *Mathematica* commands:

InitGroup[e];
Define[x^3 ,e]
Define[y^6, e]
Define[y.x, x.x.y.y]
y.y.x

Mathematica would blindly make the following "simplifications"

$$y \cdot y \cdot x \to y \cdot x \cdot x \cdot y \cdot y \to x \cdot x \cdot y \cdot y \cdot x \cdot y \cdot y \to x \cdot x \cdot y \cdot x \cdot x \cdot y \cdot y \cdot y \cdot y \to \cdots$$

indicated by the **Define** statements, creating longer and longer expressions and never stopping. The problem is not that the group does not exist; in fact problem 8.28 asks you to find a group of order 24 for which there are elements x and y such that $x^3 = e$, $y^6 = e$, and $y \cdot x = x^2 \cdot y^2$. The above infinite loop stems from trying to define this group in terms of subgroups that are not normal subgroups. Whenever we use a polycyclic series to define a group in *Mathematica* or GAP this type of infinite loop will never happen.

PROPOSITION 8.4
Let G be a finite solvable group, and let

$$G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_n = \{e\}$$

be a polycyclic series for G. If the group is defined in Mathematica *or GAP using n generators and the procedure described above, then* Mathematica *or GAP will simplify any combination of generators to a point where no further reductions are possible.*

PROOF This is not really a proof about *Mathematica* or GAP, but about the structure of polycyclic groups. However, the proposition can best be stated in terms of how *Mathematica* handles the elements of the group.

First consider the case where $n = 1$. The group G is will then be a cyclic group, say of order $m > 1$. The only **Define** statement would replace g_1^m with e, so each substitution would reduce the number of g's in the expression, and hence would eventually come to the point where no more substitutions are possible.

We can now proceed by induction on the length of the polycyclic series of G. That is, we will assume that the proposition is true for all groups with shorter polycyclic series, in particular, G_1.

Since G_0/G_1 is cyclic, we will let $u \cdot G_1$ be a generator, and let $m = |G_0/G_1|$. We will then let g_1 be one element from $u \cdot G_1$. Since g_1^m is in G_1, by induction we can let $g_1^m = b$, where b is defined in terms of the generators $\{g_2, g_3, \cdots, g_n\}$. Also, $g_1^{-1} \cdot g_i \cdot g_1$ is in G_1 for each of these generators, and so we can define $k_i = g_1^{-1} \cdot g_i \cdot g_1$ for $i = 2, 3, \ldots, n$ in terms of the generators $\{g_2, g_3, \cdots, g_n\}$. We then have the additional n **Define** commands:

Define[g_1^**m , b**]
Define[$g_2 \cdot g_1$**,** $g_1 \cdot k_2$]
Define[$g_3 \cdot g_1$**,** $g_2 \cdot k_3$]
. . . . . .
Define[$g_n \cdot g_1$**,** $g_1 \cdot k_n$]

We will call these n new **Define** commands "first category substitutions," and all previously defined definitions as "second category substitutions." Certainly these definitions are compatible with the group structure of G, so if we can simplify every combination to a unique form, this form will be the correct representation of the element.

The only thing that would go wrong is if there was some expression for which there existed an infinite sequence of substitutions from either category. Suppose that this was the case. That is, suppose we have an infinite sequence of expressions

$$u_1, u_2, u_3, \ldots$$

where each expression u_i is formed from a substitution of either of the two categories applied to u_{i-1}. Note that the u_i's do not represent elements of G, but rather expressions that are products of the generators $\{g_1, g_2, \ldots, g_n\}$. In fact, all of the u_i's are different ways of expressing the same element of G. If such an infinite sequence of expressions existed, the computer would have the potential of running into an infinite loop.

Let d represent the number of times that g_1 appears in the expression u_1. Note that if $d = 0$, then the u_1 is expressed in terms of the generators $\{g_2, g_3, \ldots, g_n\}$ of G_1. But by induction, G_1 does not form any such infinite sequences. Thus, we may assume that there is at least one occurrence of g_1 in the expression u_1. By the same argument, we can suppose that there is at least one occurrence of g_1 in all of the expressions u_i.

Consider the first appearance of the generator g_1 in each expression u_i. If we let v_i be the part of the expression occurring before this first g_1, and let w_i represent the part of the expression occurring after it, we can express u_i as $v_i \cdot g_1 \cdot w_i$. Note that v_i and w_i may be empty expressions.

Since v_1 contains no g_1's, it is in G_1 and so by our induction hypothesis, there is only a finite number of expressions that could be produced using substitutions from the second category. Let s denote the number of generators in the longest such expression.

We now will show, using induction on the number d, that an infinite sequence of substitutions is impossible. That is, we will assume that an expression with only $d-1$ occurrences of g_1 could not appear in an infinite loop. Note

that we are already using an induction hypothesis, so this is an "induction inside of an induction." We will keep the two induction arguments straight by referring to them as the "inner induction" and the "outer induction."

Notice that the first substitution of the first category,

Define[g_1ˆ**m , b**]

reduces the number of g_1's by m. All other substitutions of the first category preserve the number of g_1's while all substitutions of the second category do not affect any of the g_1's. Thus, if g_1^m is ever replaced by b, the resulting expression would have only $d-m$ occurrences of g_1, and by the inner induction hypothesis would not get into an infinite loop. Hence we can suppose that the number of g_1's that appears in any of the expressions u_i is the same, which is d.

For each expression $v_i \cdot g_1 \cdot w_i$, there are three types of substitutions that can be done:

1. A substitution of the second category applied to v_i.

2. A substitution of either category applied to w_i.

3. A substitution of the first category applied to the last generator of v_i and the first occurrence of g_1. The resulting v_{i+1} will be shorter than v_i by one symbol.

By the outer induction hypothesis, since v_i is in G_1, only a finite number of substitutions of the first type can be done before doing one of the third. Likewise, by the inside induction hypothesis, since w_i contains only $(d-1)$ occurrences of g_1, only a finite number of substitutions of the second type can be done before performing before one of type 3. But the size of v_i goes down by one each time the third type of substitution occurs, which could happen only s times. Thus, the computer will not go into an infinite loop when the generator g_1 appears d times. Thus, by the inner induction, the computer will not go into an infinite loop making substitutions on any combination of generators in $\{g_1, g_2, g_3, \ldots, g_n\}$.

We now can close the outer induction argument. Since we have shown that there cannot be an infinite number of substitutions on a combination of generators in G_0 provided that the same was true for G_1, and that G_0/G_1 was cyclic, we can see by induction that no such infinite number of substitutions is possible on the original group G. □

Because this result is the foundation that allows this set of notebooks to exist in *Mathematica* or GAP, it is included here. It gives a good example of how the tools that we have learned throughout the course, such as induction and reductio ad absurdum, can be applied consecutively to solve harder problems.

8.4 Solving the Pyraminx™

In section 2.3, we introduced a very large group called the Pyraminx™ group, formed from the different actions that can be performed on the puzzle in figure 2.3.

This group was described by four generators, r, l, b, and f, which rotated the right, left, back, or front corners 120° clockwise. The size of the group (933120 elements) makes it infeasible to list the elements in either *Mathematica* or GAP, but we still can use the tools we have learned to analyze this group. Does the group has a nontrivial center? Notice that the four corner pieces will never change location in the puzzle. The sequence of moves

ResetPuzzle
RotatePuzzle[f.r.f.r.r.f.r.f.r.r]

rotates one of these corner pieces, returning all other pieces to their original positions. It is clear that this sequence would commute with all other sequences performed on the puzzle. Since the four corners act independently, we would find at least $3^4 = 81$ elements in the center of the group. Let us call this subgroup K.

Are there elements in the center besides those in K? The sequence

ResetPuzzle
RotatePuzzle[l.l.b.f.l.l.b.f.l.l.b.f]

returns the four corner pieces to their place, while putting all the edge pieces in the right position, but reversed. If a further sequence of moves was performed from this position rather than the original position, the difference in the end positions would be that all six edges would be reversed. Thus, the above sequence of order 2 will commute with all other elements of the group. It is clear that there can be no more elements in the center, for such an element would have to keep the edge pieces in place. Hence, the center is a normal subgroup isomorphic to the group $Z_2 \times Z_3 \times Z_3 \times Z_3 \times Z_3$.

Suppose we consider the subgroup E of actions that return all of the *corners* to their original place. If x is an element of E, and y is a general element, say y rotates the front corner n degrees. Then $y \cdot x \cdot y^{-1}$ rotates the front corner $n + 0 + (-n) = 0$ degrees, so the front corner would return to its original position. Since the same is true for the other three corners, we see that E is a normal subgroup.

The intersection of E and K would be the only element that leaves both the edges and the corners fixed, the identity element. Since both E and K are normal (since K is in the center), by the direct product theorem, $E \cdot K$ is isomorphic to $E \times K$. Yet any action on the Pyraminx™ can be performed by first moving all of the edge pieces, and then moving all of the corners. Thus,

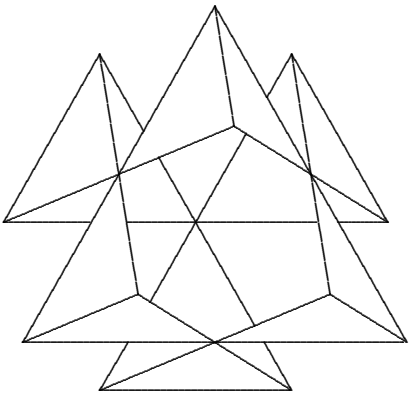


FIGURE 8.3: The Pyraminx$^{\text{TM}}$ without the corners

the entire group is in $E \cdot K$, and so the Pyraminx$^{\text{TM}}$ group is isomorphic to

$$E \times K \approx E \times Z_3 \times Z_3 \times Z_3 \times Z_3.$$

To find the structure of the subgroup E, we analyze the puzzle without the corners, as in figure 8.4 created by *Mathematica*'s **HideCorners** command.

Since there are only 12 triangles remaining, it is clear that each action could be described as a permutation of the 12 triangles. In fact, notice that turning one corner 120° moves 6 triangles—two sets of 3 triangles rotate places. Thus, each turn produces an *even* permutation of the 12 triangles, so E is a subgroup of A_{12}.

Let us now try to find a normal subgroup of E. What if we considered the subgroup of actions that returns the edge pieces to their place, but may reverse some of them? Let us call this subgroup H. Let x be an element of H, and y an element of E. The action $y^{-1}x \cdot y$ may temporarily move an edge piece out of position, but will return it to its proper place after possibly flipping it. Therefore, H will be a normal subgroup of E.

Let us determine the structure of H. At first one might think that each edge piece can be reversed independently of all of the others, but this is not true. An action that reverses only *one* edge piece would be an *odd* permutation of the triangles. So every element of H must reverse an even number of edge pieces. The sequence of moves

ResetPuzzle
RotatePuzzle[**l.f.l.b.l.b.f.b.f**]

reverses the two front edge pieces, hence it is possible to reverse two edge pieces when they are touching. Using routines like this one, we can reverse any combination of edges as long as the number of edges reversed is even.

How many elements of H will there be? If we had considered the edge pieces to be reversed independently, there would have been $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$

elements. Of these 64 possibilities, half of them reverse an even number of edges. By noticing that all elements of H besides the identity are of order 2, we find that the 32 elements of H are isomorphic to $Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2$. The quotient group E/H can now be visualized by ignoring whether the six edge pieces are reversed. Certainly this would be a subgroup of the permutations of the six edges. But again we can only consider even permutations, for the edges are moved three at a time. Thus E/H must be isomorphic to a subgroup of A_6. It is fairly clear that we can position four of the six edges in any position, so $E/H \approx A_6$.

Is E isomorphic to a semi-direct product of H with A_6? To see that it is, we need to find a copy of A_6 inside of E that contains no elements of H besides the identity. Such a subgroup is generated by the three actions

RotatePuzzle[f]
RotatePuzzle[b]
RotatePuzzle[r.f.f.r.r.f]

so the group K generated by these three sequences is isomorphic to A_6. Since it is impossible to reverse any edges with the elements of K, the intersection of K and H is the identity. Every arrangement of the edges can be obtained by first putting all of the edges into position, and then reversing several edges. Thus, $E = K \cdot H$. Therefore by the semi-direct product theorem (6.3), E is isomorphic to a semi-direct product of H with K. If we let ϕ represent the homomorphism from K to $\mathrm{Aut}(H)$, we have that

$$E \approx A_6 \ltimes_{\phi} (Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2).$$

Surprisingly, there is only one semi-direct product of this form! Let's sketch a proof of this remarkable statement.

We begin by finding all nontrivial homomorphisms from A_6 to the group $G = \mathrm{Aut}(Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2)$. The kernel of such a homomorphism would have to be a normal subgroup of A_6. But A_6 is simple, so the kernel must be just the identity. Thus the homomorphism is an isomorphism from A_6 onto a copy of A_6 in G. Let us look for copies of A_6 within the group G.

Although the group G is huge (9,999,360 elements), there are some shortcuts to this process. Consider the single element of G given by f, where

$$f(A) = B, \qquad f(B) = C, \qquad f(C) = D, \qquad f(D) = E, \quad \text{and} \quad f(E) = A,$$

and A, B, C, D, and E are five generators of the group $Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2$. The element f is of order 5, and using *Mathematica*, we can find that there are exactly 15 elements of G that commute with f. These 15 elements form a cyclic group that is generated by the element g, where

$$g(A) = A \cdot C \cdot E,$$
$$g(B) = A \cdot B \cdot D,$$

$$g(C) = B \cdot C \cdot E,$$
$$g(D) = A \cdot C \cdot D,$$
$$g(E) = B \cdot D \cdot E.$$

Notice that $g^3 = f$, and hence g commutes with f. By lemma 7.2, the number of elements of G that are conjugate to f is 9,999,360/15 = 666,624. All of these elements would be of order 5, so there are at least 666,624 elements of G of order 5. By the second Sylow theorem (7.4), all 5-Sylow subgroups of G are conjugate. Thus each 5-Sylow subgroup would contain 1, 2, or 4 elements conjugate to f. But the third Sylow theorem (7.5) eliminates the first two possibilities. Therefore, all 666,624 elements of G of order 5 are conjugate.

For each of these elements of order 5, let us determine the number of copies of A_6 in G that contain that element. Because the elements of order 5 all conjugate, we only need to consider the number of copies of A_6 in G that contain the element f. Since A_6 is generated by (12345) and (13)(46), it is logical to look for elements in G that are of order 2, and that together with f generate a copy of A_6.

Mathematica can find exactly 6975 elements of G of order 2. Notice that $(12345) \cdot (13)(46) = (1465)(23)$, which is of order 4, and $(12345) \cdot (12345) \cdot (13)(46) = (15246)$, which is of order 5. Thus, to determine which of these elements of G could correspond to the element (13)(46), we need to find the elements μ of G such that $f \cdot \mu$ is of order 4, and $f \cdot f \cdot \mu$ is of order 5. By searching though the 6975 elements, *Mathematica* found exactly 90 such elements. Each of these 90 elements, together with f, generated a copy of A_6. However, each copy of A_6 contained 10 of the 90 elements. Thus, *Mathematica* came up with nine copies of A_6 in G that contain the element f.

Even though there may be many other copies of A_6 in G, all copies must contain an element of order 5, and we already mentioned that all such elements would be conjugate to f in G. Proposition 6.7 tells us that two semi-direct products are isomorphic if the images of the ϕ's are conjugate. Thus, we may assume that the image of ϕ is one of the nine copies of A_6 in G that contain f, which we will call H. But notice however $g^{-1} \cdot H \cdot g$ and $g^{-2} \cdot H \cdot g^{-2}$ would also be copies of A_6 containing the element f, and H cannot be the same subgroup as $g^{-1} \cdot H \cdot g$, since this would imply that A_6 has an automorphism of order 15, which is not true. Thus, the nine copies of A_3 in G containing the element f appear as three collections of three subgroups, with the three subgroups in each collection being conjugate to one another. Therefore, by proposition 6.7 there are only three semi-direct products we will have to consider.

Because these groups are insoluble, these semi-direct products must be represented using 5×5 matrices instead of using generators. In all three cases, the orders of the elements are given in table 8.2.

Although this gives some strong evidence that the three possible semi-direct products are in fact isomorphic to each other, the actual isomorphisms had to be verified by *Mathematica*. Therefore, there is only one semi-direct product of A_6 and $Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2$. We then can describe the Pyraminx$^{\text{TM}}$

TABLE 8.2: Orders of $A_6 \ltimes (Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2)$

1	element of order 1,
391	elements of order 2,
800	elements of order 3,
2520	elements of order 4,
2304	elements of order 5,
1760	elements of order 6,
1440	elements of order 8,
2304	elements of order 10,
11520	elements total.

group as being the group isomorphic to

$$(A_6 \ltimes (Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2)) \times Z_3 \times Z_3 \times Z_3 \times Z_3.$$

Knowing the structure of the group allows us the solve the puzzle! Here is the strategy based on this decomposition of the group.

1. First put all of the edge pieces in place. We can begin with the bottom, then rotate the front and back corners until the back two edges are in the right place (they may be reversed). Finally, rotate the front corner until all six edges are in place.

2. At this point, an even number of edges will be reversed. We can find routines that will flip two, four, or six of the edges. These may rotate corners in the process.

3. Now only the four corner pieces are out of position. We can find routines to rotate these into position.

To find a combination of the four moves f, b, r, and l that will accomplish these goals, we can have GAP help us. First we can number the 24 triangles, as in figure 8.4. Then the permutation (4 23 14)(5 24 15)(6 19 16) can represent r, $l = (8\ 16\ 21)(9\ 17\ 22)(10\ 18\ 23)$, $f = (1\ 13\ 7)(2\ 14\ 8)(6\ 18\ 12)$, and finally $b = (2\ 10\ 19)(3\ 11\ 20)(4\ 12\ 21)$. We can then enter the Pyraminx$^{\text{TM}}$ group as a subgroup of S_{24}.

```
gap> r := (4,23,14)(5,24,15)(6,19,16);
(4,23,14)(5,24,15)(6,19,16)
gap> l := (8,16,21)(9,17,22)(10,18,23);
(8,16,21)(9,17,22)(10,18,23)
gap> f := (1,13,7)(2,14,8)(6,18,12);
(1,13,7)(2,14,8)(6,18,12)
gap> b := (2,10,19)(3,11,20)(4,12,21);
(2,10,19)(3,11,20)(4,12,21)
gap> p := Group(r,l,f,b);
<permutation group with 4 generators>
gap> Size(p);
933120
```

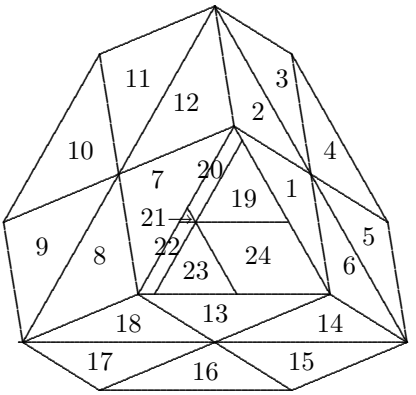


FIGURE 8.4: The Pyraminx$^{\text{TM}}$ with numbered faces

Now that we have the group representing the puzzle entered into GAP, The natural question is how to express any given permutation in this group in terms of f, b, r, and l in the most efficient way. For example, suppose we want to find an efficient way to rotate just the right corner piece clockwise, that is, the permutation (5,24,15). Here is how we can do it:

```
gap> phi:=EpimorphismFromFreeGroup(p:names:=["r","l","f","b"]);
[ r, l, f, b ] -> [ (4,23,14)(5,24,15)(6,19,16),
  (8,16,21)(9,17,22)(10,18,23), (1,13,7)(2,14,8)(6,18,12),
  (2,10,19)(3,11,20)(4,12,21) ]
gap> PreImagesRepresentative( phi, (5,24,15) );
r*b*r^-2*b^-1*r*b*r*b^-1
```

This creates a homomorphism from the group generated by the letters f, b, r, and l to the permutation group of the puzzle. By finding the inverse homomorphism of a permutation, we get a sequence of letters, which tells us how to accomplish this task on the puzzle. This particular task of rotating the corner piece, and leaving everything else fixed, is done in eight moves.

```
r*b*r*b^-1*r*b*r*b^-1;
(5,24,15)
```

In flipping edges, we have the advantage that we do not care if corners are rotated in the process. So we can enter versions of r, l, f, and b that ignore the corner pieces. For example, to flip the top and front left edges, we look for the permutation $(2,12)(8,18)$.

```
gap> r := (4,23,14)(6,19,16);;
gap> l := (8,16,21)(10,18,23);;
gap> f := (2,14,8)(6,18,12);;
gap> b := (2,10,19)(4,12,21);;
gap> p := Group(r,l,f,b);;
gap> Size(p);
```

```
11520
gap> phi:=EpimorphismFromFreeGroup(p:names:=["r","l","f","b"]);;
gap> PreImagesRepresentative(phi, (2,12)(8,18));
r*l^-1*b^-1*l*r^-1*f^-1
gap> r*l^-1*b^-1*l*r^-1*f^-1
(2,12)(8,18)
gap> PreImagesRepresentative(phi, (6,14)(10,21));
r^-1*b*l*b^-1*l^-1*r^-1*b^-1*r^-1
gap> r*b*r*l*b*l^-1*b^-1*r;
(6,14)(10,21)
```

Note that in the last example, we took the inverse of the combination that GAP gave us to produce a simpler looking combination. We summarize the necessary moves in tables 8.3 and 8.4.

TABLE 8.3: Flipping edges into position

$l^{-1} \cdot b \cdot f \cdot l^{-1} \cdot b \cdot f \cdot l^{-1} \cdot b \cdot f$	flip all six edges
$f \cdot b \cdot r^{-1} \cdot l \cdot r \cdot b^{-1}$	flip two front edges
$b \cdot l \cdot b \cdot r \cdot l \cdot r^{-1} \cdot l^{-1} \cdot b$	flip top & bottom edges
$f \cdot r \cdot l^{-1} \cdot b \cdot l \cdot r^{-1}$	flip top & front left edges
$r \cdot l^{-1} \cdot b \cdot l \cdot r^{-1} \cdot f$	flip top & front right edges
$r \cdot b \cdot r \cdot l \cdot b \cdot l^{-1} \cdot b^{-1} \cdot r$	flip left rear & front right edges
$l \cdot r \cdot l \cdot b \cdot r \cdot b^{-1} \cdot r^{-1} \cdot l$	flip right rear & left front edges
$r \cdot b \cdot l^{-1} \cdot f \cdot l \cdot b^{-1}$	flip bottom & front right edges
$l \cdot b \cdot f^{-1} \cdot r \cdot f \cdot b^{-1}$	flip bottom & front left edges
$b \cdot r \cdot f^{-1} \cdot l \cdot f \cdot r^{-1}$	flip top & left rear edges
$b \cdot l \cdot r^{-1} \cdot f \cdot r \cdot l^{-1}$	flip top & right rear edges
$b \cdot f \cdot l^{-1} \cdot r \cdot l \cdot f^{-1}$	flip rear two edges
$l \cdot f \cdot r^{-1} \cdot b \cdot r \cdot f^{-1}$	flip bottom & left rear edges
$r \cdot f \cdot b^{-1} \cdot l \cdot b \cdot f^{-1}$	flip bottom & right rear edges
$l \cdot r \cdot b^{-1} \cdot f \cdot b \cdot r^{-1}$	flip two left hand edges
$r \cdot l \cdot f^{-1} \cdot b \cdot f \cdot l^{-1}$	flip two right hand edges

TABLE 8.4: Rotating corners into position

$f \cdot r \cdot f \cdot r^{-1} \cdot f \cdot r \cdot f \cdot r^{-1}$	rotate front corner 120° clockwise
$l \cdot r \cdot l \cdot r^{-1} \cdot l \cdot r \cdot l \cdot r^{-1}$	rotate left corner 120° clockwise
$r \cdot b \cdot r \cdot b^{-1} \cdot r \cdot b \cdot r \cdot b^{-1}$	rotate right corner 120° clockwise
$b \cdot r \cdot b \cdot r^{-1} \cdot b \cdot r \cdot b \cdot r^{-1}$	rotate back corner 120° clockwise

By applying these four routines once or twice, we can get all four corners into position, and have solved the puzzle!

Notice that our three steps can be expressed in terms of a subnormal series for the Pyraminx$^{\mathrm{TM}}$ group:

$$(A_6 \ltimes (Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2)) \times Z_3 \times Z_3 \times Z_3 \times Z_3 \supset$$
$$Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_3 \times Z_3 \times Z_3 \times Z_3 \supset Z_3 \times Z_3 \times Z_3 \times Z_3 \supset \{e\}.$$

This same type of analysis can be used to solve other puzzles, such as the Rubik's Cube®. Several problems in the homework relate to this puzzle. Thus, we can see a practical application of the properties of groups that we have studied throughout the course.

Problems for Chapter 8

Interactive Problems

8.1 Use *Mathematica* or GAP to find the derived series of the group Q:

```
InitGroup[e];
Define[i^4, e]; Define[j^2, i^2]
Define[j.i, i.i.i.j]
Define[i^(-1), i^3]; Define[j^(-1), i.i.j]
Q = Group[{i, j}]
```

or, in GAP,

```
gap> f := FreeGroup("i","j");; i := f.1;; j := f.2;;
gap> Q := f/[i^4,j^2/(i^2),j*i/(i^3*j)];; i := Q.1;; j := Q.2;;
```

Add any subgroups necessary to make this series a composition series.

8.2 Use *Mathematica*'s **Commutator** or GAP's `CommutatorSubgroup` command as an alternative way to show that $\mathrm{Aut}(Z_{24}^*)$ is insoluble. Load this group with the commands

```
InitPermMultiplication
A = Group[{149, 735}]
```

or

```
gap> A := Group( (1,2,3)(4,6,5), (2,4)(6,7) );
Group([ (1,2,3)(4,6,5), (2,4)(6,7) ])
```

and find A'. Note that both *Mathematica* and GAP can find the derived group quickly.

8.3 Find the derived group series of the following group:

```
InitPermMultiplication
G = Group[{6782, 10159}]
```

```
gap> G := Group( NthPerm(6782), NthPerm(10159) );
Group([ (1,6,4,2)(3,8,7,5), (2,3,5)(6,7,8) ])
```

TABLE 8.5: Mystery group B used in problem 8.5

·	1	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	1	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
I	I	L	K	N	M	1	O	J	Q	T	S	V	U	P	W	R
J	J	O	L	I	N	K	1	M	R	W	T	Q	V	S	P	U
K	K	J	M	L	O	N	I	1	S	R	U	T	W	V	Q	P
L	L	M	N	O	1	I	J	K	T	U	V	W	P	Q	R	S
M	M	1	O	J	I	L	K	N	U	P	W	R	Q	T	S	V
N	N	K	1	M	J	O	L	I	V	S	P	U	R	W	T	Q
O	O	N	I	1	K	J	M	L	W	V	Q	P	S	R	U	T
P	P	Q	R	S	T	U	V	W	L	M	N	O	1	I	J	K
Q	Q	T	S	V	U	P	W	R	M	1	O	J	I	L	K	N
R	R	W	T	Q	V	S	P	U	N	K	1	M	J	O	L	I
S	S	R	U	T	W	V	Q	P	O	N	I	1	K	J	M	L
T	T	U	V	W	P	Q	R	S	1	I	J	K	L	M	N	O
U	U	P	W	R	Q	T	S	V	I	L	K	N	M	1	O	J
V	V	S	P	U	R	W	T	Q	J	O	L	I	N	K	1	M
W	W	V	Q	P	S	R	U	T	K	J	M	L	O	N	I	1

What group is G' isomorphic to? Is G a semi-direct product of two familiar groups?

8.4 Use a polycyclic series of A_4 to enter this group into GAP or *Mathematica*.

8.5 Find a polycyclic series of group B of order 16 given in table 8.5, and use this to enter the group into GAP or *Mathematica*.

8.6 Find a polycyclic series of group C of order 16 given in table 8.6, and use this to enter the group into GAP or *Mathematica*.

8.7 Find a polycyclic series of group D of order 16 given in table 8.7, and use this to enter the group into GAP or *Mathematica*.

Non-Interactive Problems

8.8 Show that any group of order p^n, where p is prime, is solvable.
Hint: See corollary 7.2.

8.9 Let

$$G = Z_{12} \supseteq A_1 = \{0, 3, 6, 9\} \supseteq \{0\}$$

TABLE 8.6: Mystery group C used in problem 8.6

·	1	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	1	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
F	F	1	H	G	J	I	L	K	N	M	P	O	R	Q	T	S
G	G	H	1	F	K	L	I	J	O	P	M	N	S	T	Q	R
H	H	G	F	1	L	K	J	I	P	O	N	M	T	S	R	Q
I	I	K	J	L	M	O	N	P	Q	S	R	T	1	G	F	H
J	J	L	I	K	N	P	M	O	R	T	Q	S	F	H	1	G
K	K	I	L	J	O	M	P	N	S	Q	T	R	G	1	H	F
L	L	J	K	I	P	N	O	M	T	R	S	Q	H	F	G	1
M	M	N	O	P	Q	R	S	T	1	F	G	H	I	J	K	L
N	N	M	P	O	R	Q	T	S	F	1	H	G	J	I	L	K
O	O	P	M	N	S	T	Q	R	G	H	1	F	K	L	I	J
P	P	O	N	M	T	S	R	Q	H	G	F	1	L	K	J	I
Q	Q	S	R	T	1	G	F	H	I	K	J	L	M	O	N	P
R	R	T	Q	S	F	H	1	G	J	L	I	K	N	P	M	O
S	S	Q	T	R	G	1	H	F	K	I	L	J	O	M	P	N
T	T	R	S	Q	H	F	G	1	L	J	K	I	P	N	O	M

TABLE 8.7: Mystery group D used in problem 8.7

·	1	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	1	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
L	L	M	N	O	P	Q	R	1	T	U	V	W	X	Y	Z	S
M	M	N	O	P	Q	R	1	L	U	V	W	X	Y	Z	S	T
N	N	O	P	Q	R	1	L	M	V	W	X	Y	Z	S	T	U
O	O	P	Q	R	1	L	M	N	W	X	Y	Z	S	T	U	V
P	P	Q	R	1	L	M	N	O	X	Y	Z	S	T	U	V	W
Q	Q	R	1	L	M	N	O	P	Y	Z	S	T	U	V	W	X
R	R	1	L	M	N	O	P	Q	Z	S	T	U	V	W	X	Y
S	S	Z	Y	X	W	V	U	T	O	N	M	L	1	R	Q	P
T	T	S	Z	Y	X	W	V	U	P	O	N	M	L	1	R	Q
U	U	T	S	Z	Y	X	W	V	Q	P	O	N	M	L	1	R
V	V	U	T	S	Z	Y	X	W	R	Q	P	O	N	M	L	1
W	W	V	U	T	S	Z	Y	X	1	R	Q	P	O	N	M	L
X	X	W	V	U	T	S	Z	Y	L	1	R	Q	P	O	N	M
Y	Y	X	W	V	U	T	S	Z	M	L	1	R	Q	P	O	N
Z	Z	Y	X	W	V	U	T	S	N	M	L	1	R	Q	P	O

and

$$G = Z_{12} \supseteq B_1 = \{0, 2, 4, 6, 8, 10\} \supseteq B_2 = \{0, 4, 8\} \supseteq \{0\}$$

be two subnormal series for Z_{12}. Find all of the subgroups shown in figure 8.2, and show that the quotient groups indicated by the arrows are indeed isomorphic.

For problems **8.10** through **8.18**: Write out a composition series for the group.

8.10 Z_{15}^*	**8.13** $Z_{12} \times Z_{18}$	**8.16** D_5
8.11 Z_{24}^*	**8.14** The quaternion group Q	**8.17** D_6
8.12 Z_{21}^*	**8.15** D_4	**8.18** S_6

8.19 Show that there are exactly three possible composition series for A_4.

8.20 Show that S_n is solvable for $n < 5$, but is insoluble for $n > 4$.

8.21 Find an example of two non-isomorphic groups for which the composition factors are isomorphic.

8.22 Find two groups of the same order with composition series of different lengths.

8.23 Find a non-simple group for which all of the composition factors are non-cyclic.

8.24 Show that $[z \cdot x \cdot z^{-1}, z \cdot y \cdot z^{-1}] = z \cdot [x, y] \cdot z^{-1}$.

8.25 Let G be the group from example 1.4 in section 1.4, the group of *linear functions* of the form $f(x) = mx + b$, with $m, b \in \mathbb{R}$, $m \neq 0$. By finding the derived group G', show that this group is solvable.

8.26 Show that if G is a non-cyclic simple group, then $G' = G$. Is it true that if $G' = G$, then G must be simple?

8.27 Throughout this course, we have encountered a number of groups of order 16. Here is a list of some of these groups:

$$Z_{16}, \qquad Z_8 \times Z_2, \qquad Z_4 \times Z_4, \qquad Z_4 \times Z_2 \times Z_2, \qquad Z_2 \times Z_2 \times Z_2 \times Z_2,$$

three groups of the form $Z_2 \ltimes_\phi Z_8$ in section 6.4 (one is D_8),

$$Z_2 \times Q, \qquad Z_2 \times D_4, \qquad Z_4 \ltimes Z_4 \text{ studied in this chapter,}$$

and three mystery groups B, C, and D found in problems 8.5, 8.6, and 8.7. Show that these 14 groups are all non-isomorphic. (In fact, these are all of the non-isomorphic groups of order 16.)

Hint: Find the number of elements of order 2 in each of the 14 groups. Note that group B has only 1's and L's along its diagonal, whereas group C has three different elements along its diagonal.

8.28 Show that there is a group of order 24 for which there are two elements x and y that generate the group such that $x^3 = y^6 = e$, and $y \cdot x = x^2 \cdot y^2$.

Hint: What are the orders of the elements $x \cdot y$ and $y \cdot x$? Determine the subgroup generated by these two elements.

8.29 Let G be an infinite group such that every element besides the identity has order 2. Show that G is solvable, yet G does not have a polycyclic series.

8.30 Let H and K be two subgroups of G. Prove that the mutual commutator $[H, K]$ is a normal subgroup of the group generated by the elements of H and K.

For problems **8.31** through **8.33**, find the derived series of the group.

8.31 D_4 **8.32** D_5 **8.33** The quaternion group Q

8.34 If G is a group, define the sequence $G_1 = [G, G]$, $G_2 = [G, G_1]$, $G_3 = [G, G_2], \ldots$. G is said to be *nilpotent* if $|G_n| = 1$ for some n. Prove that if G is nilpotent, then G is solvable.

Hint: Prove that G_n contains the n-th derived group of G.

8.35 Find a solvable group that is not nilpotent. (See problem 8.34.)

8.36 Show that a group of order p^n, where p is prime, is nilpotent. (See problem 8.34 and corollary 7.2.)

8.37 Prove that if the refinement theorem (8.1) is applied to two *normal* series, the resulting series will be normal. That is, if A_u and B_v are such that

$$G = A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n = \{e\},$$

and

$$G = B_0 \supseteq B_1 \supseteq B_2 \supseteq \cdots \supseteq B_m = \{e\},$$

where each A_i and B_j is a normal subgroup of G (not just the previous group), then the $A_{i,j}$ and $B_{j,i}$ given by the refinement theorem will all be normal subgroups of G.

Hint: Use the result of problem 4.21.

8.38 A *chief* series is a normal series for which no refinements produce normal series. Show that the Jordan-Hölder theorem (8.2) applies to chief series as well as to composition series. That is, show that if

$$G = A_0 \supseteq A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n = \{e\}$$

and

$$G = B_0 \supseteq B_1 \supseteq B_2 \supseteq \cdots \supseteq B_m = \{e\}$$

are two chief series, then $n = m$, and the quotient groups of the first series are isomorphic to the quotient groups of the second in some order. (Use the result from problem 8.37.)

8.39 A group is called *supersolvable* if there is a chief series with cyclic factors. Show that if G is supersolvable, then G' is nilpotent. (See problems 8.34 and 8.38.)

8.40 Using the orders of the subgroup E of the Pyraminx™ group given in the chapter, determine the number of elements of the Pyraminx™ group that are of order 1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 24, and 30. Verify that the sum of these numbers totals 933,120.

8.41 Consider a $2 \times 2 \times 2$ Rubik's Cube®, consisting of just eight corner pieces. Determine the size of the group of actions on this cube. Express the group of actions as a semi-direct product of two familiar groups. You do not need to show that this semi-direct product is unique.

Hint: It is impossible to rotate just one corner, and leave the others in place. Is it possible to move just two of the corners?

8.42 Consider a standard Rubik's Cube®. What is the size of the group of actions? What is the center of this group?

Chapter 9

Introduction to Rings

9.1 Groups with an Additional Operation

Many of the groups studied in the previous chapters possessed some additional structure. From now on, we will consider those groups that have not just one, but two operations defined on the set of elements. In other words, not only will we be able to multiply elements together as we did for groups, but we also will be able to *add* elements together.

The simplest example to consider is the group of integers, $\mathbb{Z}$. This is a group under addition, but we can also multiply two integers together. This extra operation gives $\mathbb{Z}$ a richer structure than standard groups.

Subgroups of $\mathbb{Z}$ should also be considered. A typical example would be the set of even integers. Once again, we have both addition and multiplication defined on this set, since both the sum and the product of two even integers yield even integers.

Another example of a group possessing two operations is the group of all rational numbers $\mathbb{Q}$ of the form p/q, where p is an integer and q is a positive integer. Although $\mathbb{Q}$ is an abelian group under addition, it is almost a group under multiplication as well. The multiplicative inverse exists for all elements except 0. If we consider the remaining elements $\mathbb{Q} - \{0\}$, denoted $\mathbb{Q}^*$, we have a multiplicative group.

One way to illustrate the rationals graphically can be seen by executing the command

ShowRationals[–5, 5]

which draws figure 9.1. This figure helps to visualize the rational numbers from -5 to 5 using a sequence of rows. The n-th row represents the rational numbers with denominator n when expressed in simplest form. In principle there would be an infinite number of rows, getting closer and closer to each other as they get close to the axis.

Figure 9.1 suggests the following.

PROPOSITION 9.1

If a and b are any two different real numbers, then there is a rational number between a and b.

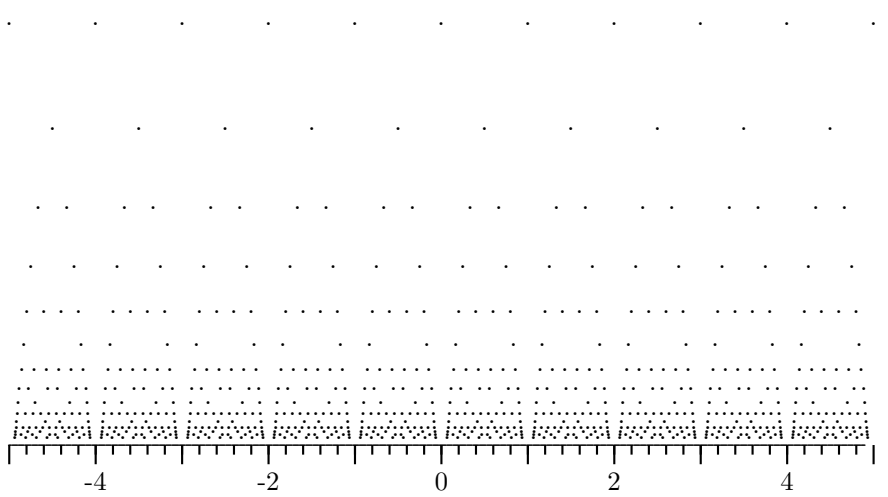


FIGURE 9.1: Plot depicting the rational numbers

PROOF Let $x = |a - b|$. Since x is not zero, we let q be any number that is greater than $1/x$. Then $|a \cdot q - b \cdot q| = q \cdot x > 1$, so there must be an integer between $a \cdot q$ and $b \cdot q$, which we will call p. But then p/q will be between a and b, and the proposition is proved. □

From this proposition, we can keep dividing the interval up into smaller and smaller pieces to show that there are in fact an infinite number of rational numbers between any two real numbers. This would make it seem that the number of rational numbers is "doubly infinite," since there are an infinite number of integers, and an infinite number of rational numbers between each pair of integers. But surprisingly, the set of rational numbers is no larger than the set of the integers. To understand what is meant by this statement, let us first show how we can compare the sizes of two infinite sets.

DEFINITION 9.1 A set S is called *countable* if there is an infinite sequence of elements from the set that includes every member of the set.

What do sequences have to do with comparing the sizes of two sets? A sequence can be considered as a function between the set of positive integers and the set S. If a sequence manages to include every member of the set S, then it stands to reason that there are at least as "many" positive integers as there are elements of S. The shocking fact is that even though it would first appear that there must be infinitely many more rational numbers than integers, in fact the two sets have the same size.

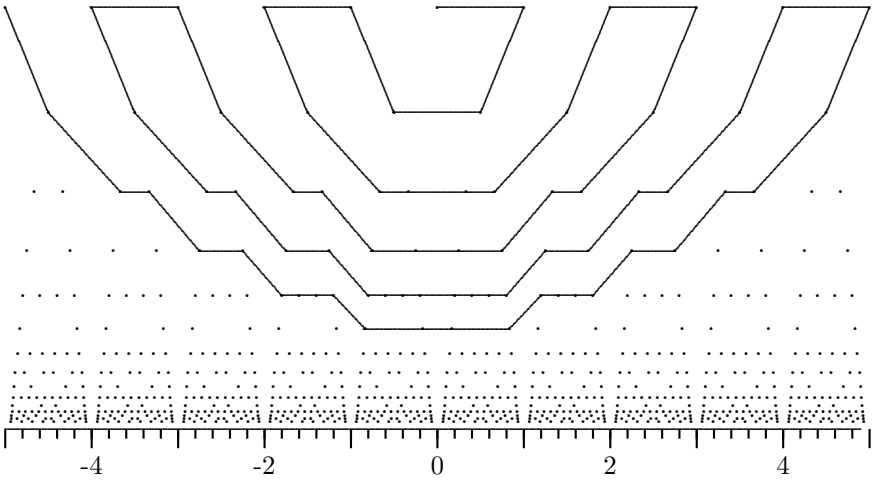


FIGURE 9.2: Beginning of a path that will hit every rational number

PROPOSITION 9.2
The set of rationals forms a countable set.

PROOF In order to show that the rationals are countable, we need a sequence that will eventually contain every rational somewhere in the sequence. Equivalently, we can connect the dots of figure 9.1 using a pattern that would, in principle, reach every dot of figure 9.1 extended to infinity. There are of course many ways to do this, but one way is given in figure 9.2. This path starts at 0, and swings back and forth, each time hitting the rationals on the next row. Since there are an infinite number of rows, we can extend this pattern indefinitely, and every rational number will eventually be hit by this path. This path gives rise to the sequence

$$\{0, 1, \frac{1}{2}, \frac{-1}{2}, -1, -2, \frac{-3}{2}, \frac{-2}{3}, \frac{-1}{3}, \frac{1}{3}, \frac{2}{3}, \frac{3}{2}, 2, 3, \ldots\}$$

which contains every rational number, so we have shown that the rationals form a countable set. □

Even though we have shown that there are an infinite number of rational numbers between any two numbers, the natural question to ask is whether there are numbers that are not rational. The first discovery of a number that was not rational was $\sqrt{2}$, proven by the Greeks. [12, p. 82]

PROPOSITION 9.3
There is no rational number p/q such that $(p/q)^2 = 2$.

PROOF Suppose that there was such a rational number, p/q. Let us further suppose that p/q is in simplest form, so that p and q are integers with no common factors. We could rewrite the equation $(p/q)^2 = 2$ as

$$p^2 = 2q^2.$$

This would indicate that p^2 is an even number, which implies that p is even.

Next, we make the substitution $p = 2r$, where r is an integer. Making this substitution for p, we get

$$(2r)^2 = 2q^2 \qquad \text{or} \qquad 2r^2 = q^2.$$

This would indicate that q^2, and hence q, is even. But this contradicts the fact that p/q was written in simplest form. Thus, there is no rational number whose square is 2. ∎

This proof is an example of a *reductio ad absurdum* proof. These types of proofs are particularly effective to prove that something is impossible.

The real numbers $\mathbb{R}$ that are not rational are called *irrational* numbers. Irrational numbers are characterized by the fact that their decimal representation never repeats.

We have already proven that there is, in essence, the same number of rational numbers as integers. This may not come as too much of a shock, since both sets are infinite, so logically two infinite sets ought to be the same size. But the set of real numbers is also infinite, so one might be tempted to think that there is the same number of real numbers as integers. However, the number of reals is "more infinite" then the number of integers. In other words, we cannot construct a sequence of real numbers that contains every real number, as we did for rational numbers. This surprising fact was proved by Georg Cantor (1845-1913) using a classic argument. [11, p. 670]

THEOREM 9.1: Cantor's Diagonalization Theorem
The set of all real numbers between 0 and 1 is uncountable. *That is, there cannot be a sequence of numbers that contains every real number between 0 and 1.*

PROOF We begin by assuming that we can form such a sequence

$$\{a_1, a_2, a_3, \ldots\}$$

and work to find a contradiction. The plan is to find a number b that cannot be in this list. We can do this by forcing b to have a different first digit than a_1, a different second digit than a_2, a different third digit than a_3, and so on. The only technical problem with this is that some numbers have two decimal representations, such as

$$0.348600000000000000\ldots = 0.348599999999999999\ldots.$$

For these numbers, all we need to do is require that *both* representations are in the list. (That is, some rational numbers will appear twice on the list with different decimal representations.)

We now can find a number b using any number of procedures, such as letting the n-th digit of b be one more than the n-th digit of a_n, modulo 10. For example, if the list of numbers is

$$\begin{aligned} a_1 &= 0.94837490123798570\ldots \\ a_2 &= 0.83840000000000000\ldots \\ a_3 &= 0.83839999999999999\ldots \\ a_4 &= 0.34281655343424444\ldots \end{aligned}$$

then $b = 0.0499\ldots$. Certainly b is missing from the list, since it differs from each member of the list by at least one digit. This contradiction proves the theorem. □

Not only do $\mathbb{Z}$, $\mathbb{Q}$, and the real numbers $\mathbb{R}$ allow for an additional operation to be defined on them but also some groups from chapter 1. Take for example the groups formed by modular arithmetic, such as Z_6.

DefSumMod[6]
MultTable[{0, 1, 2, 3, 4, 5}]

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

This multiplication table can be displayed in GAP by the command

```
gap> MultTable([0..5]);

+|0  1  2  3  4  5
-+---------------
0|0  1  2  3  4  5
1|1  2  3  4  5  0
2|2  3  4  5  0  1
3|3  4  5  0  1  2
4|4  5  0  1  2  3
5|5  0  1  2  3  4
```

A natural second operation would be multiplication modulo 6, defined by

DefMultMod[6]
MultTable[{0, 1, 2, 3, 4, 5}]

TABLE 9.1: $(\cdot)$ Mod 6

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

which produces table 9.1. In GAP, we can display this table except for the 0 row and column with

```
gap> MultTable([1..5]);

*|1  2  3  4  5
-+-------------
1|1  2  3  4  5
2|2  4  0  2  4
3|3  0  3  0  3
4|4  2  0  4  2
5|5  4  3  2  1
```

Even though this table does not possess the "Latin square" property we have seen in the group tables, the second operation need not have this familiar property.

Here is one last example of appending an additional operation on a group. The following command produces the quaternion group Q of order 8 which we studied in chapter 4:

```
gap> InitQuaternions();
#I  default 'IsGeneratorsOfMagmaWithInverses' method returns
'true' for [ i, j ]
gap> MultTable(Q);

*     |(-1)*e (-1)*i (-1)*j (-1)*k k      j      i      e
------+-----------------------------------------------------
(-1)*e|e      i      j      k      (-1)*k (-1)*j (-1)*i (-1)*e
(-1)*i|i      (-1)*e k      (-1)*j j      (-1)*k e      (-1)*i
(-1)*j|j      (-1)*k (-1)*e i      (-1)*i e      k      (-1)*j
(-1)*k|k      j      (-1)*i (-1)*e e      i      (-1)*j (-1)*k
k     |(-1)*k (-1)*j i      e      (-1)*e (-1)*i j      k
j     |(-1)*j k      e      (-1)*i i      (-1)*e (-1)*k j
i     |(-1)*i e      (-1)*k j      (-1)*j k      (-1)*e i
e     |(-1)*e (-1)*i (-1)*j (-1)*k k      j      i      e
```

The corresponding *Mathematica*® commands

InitQuaternions
Q = {1, I, J, K, −1, −I, −J, −K}
MultTable[Q]

produce table 4.3 that we have seen before. When written in this way, the quaternion elements are reminiscent of the cross product between two vectors. In fact, in order to get a second operation on this set, we can consider *adding* multiples of these elements together like vectors, forming such elements as

```
gap> (i - 2*j - k) + (3*i + j - 2*k);
(4)*i+(-1)*j*(-3)*k
```

which represents the vector $\langle 4, -1, -3\rangle$. Unfortunately, as we multiply these "vectors" together, we find elements of the form

```
gap> (i - 2*j - k) * (3*i + j - 2*k);
(-3)*e+(5)*i+(-1)*j+(7)*k
```

which would represent the *four*-dimensional vector $\langle -3, 5, -1, 7\rangle$.

PROPOSITION 9.4

The set of nonzero four-dimensional vectors forms a non-abelian group using the multiplication table for the quaternion group Q.

PROOF If

$$x = a + bi + cj + dk$$

is nonzero, then

$$x^{-1} = \frac{a}{a^2+b^2+c^2+d^2} + \frac{-b}{a^2+b^2+c^2+d^2}\, i + \frac{-c}{a^2+b^2+c^2+d^2}\, j + \frac{-d}{a^2+b^2+c^2+d^2}\, k$$

forms a multiplicative inverse, since it is a simple exercise to show that $x \cdot x^{-1} = 1$, the multiplicative identity. (See problem 9.15.) Note that since $x \neq 0$, the common denominator $a^2+b^2+c^2+d^2 > 0$. It is easy to see that multiplication is closed. The only hard part is to show that the associative law holds, which is best done via a program like *Mathematica*.

Given that the associative law holds, it is easy to see that the product of two nonzero vectors must be nonzero. If $x \cdot y = 0$, and $x \neq 0$, then

$$y = (x^{-1} \cdot x) \cdot y = x^{-1} \cdot (x \cdot y) = x^{-1} \cdot 0 = 0.$$

Thus, if both $x \neq 0$ and $y \neq 0$, then $x \cdot y \neq 0$. □

We call the group of four-dimensional vectors of the form $a+bi+cj+dk$ the *quaternions*, denoted by $\mathbb{H}$ after their discoverer, William Rowan Hamilton (1805-1865).

We have now seen several examples of groups that have additional structure in the form of a second operation. In the next section we will tie all of these examples together, discovering which properties all of the examples have in common.

9.2 The Definition of a Ring

In the preceding section we saw many examples of groups that exhibit not one but two operations defined on them. One of these operations is represented with the plus sign, and the other is usually denoted with a dot. However, some of the different groups we looked at possessed additional properties. To help us organize our findings, let us construct a checklist from table 9.2. This checklist is already started, since all six of these groups are closed under addition. Before going on, please try to complete table 9.2.

We want to pay special attention to the properties that hold for *all* of the groups studied so far. In fact, let us define a *ring* as a group possessing all of these properties. In this way, we force all six of the above groups to be rings.

DEFINITION 9.2 A *ring* is an abelian group with the operation $(+)$ on which a second associative operation $(\cdot)$ is defined such that the two distributive laws

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c)$$

and

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$$

hold for all a, b, and c in the ring.

For any ring we will use the symbol 0 to denote the additive identity of a ring, and the notation $-x$ for the additive inverse of x.

Even though we defined a ring such that all six of the groups in table 9.2 are rings, you may also have noticed that many of the groups possessed additional properties. We will give names to rings with some of these extra properties.

DEFINITION 9.3 A ring for which $x \cdot y = y \cdot x$ for all elements x and y is called a *commutative ring.*

DEFINITION 9.4 A ring for which there is an element e such that

$$x \cdot e = e \cdot x = x$$

for all elements x in the ring is called a *ring with identity.* The element e is called the *multiplicative identity* of the ring.

Using only the definition of rings, we can prove a few things that are true for all rings.

TABLE 9.2: Property checklist for several groups

Property	$\mathbb{Z}$	Even Integers	$\mathbb{Q}$	Reals	Z_6	Quaternions
Closed under Addition	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Closed under Multiplication						
$(a+b)+c=$ $a+(b+c)$						
$(a\cdot b)\cdot c=$ $a\cdot(b\cdot c)$						
Additive Identity (0)						
Multiplicative Identity (1)						
Additive Inverses Exist						
Multiplicative Inverses Exist Except for 0						
$a+b=b+a$						
$a\cdot b=b\cdot a$						
$a\cdot b=0$ only if a or $b=0$						
$(a+b)\cdot c=$ $a\cdot c+b\cdot c$						
$a\cdot(b+c)=$ $a\cdot b+a\cdot c$						

LEMMA 9.1
If x is any element in a ring, then $0 \cdot x = x \cdot 0 = 0$, where 0 is the additive identity.

PROOF This proof is just a little tricky because there are no other propositions to rely on. Thus, every step must directly use one of the nine properties of rings. (The temptation is to rely on some property we suspect is true, but haven't yet proven.)

Note that

$$(0 \cdot x + 0 \cdot x) = (0 + 0) \cdot x = 0 \cdot x,$$

so

$$(0 \cdot x + 0 \cdot x) + (-(0 \cdot x)) = 0 \cdot x + (-(0 \cdot x)) = 0.$$

Hence

$$0 \cdot x + (0 \cdot x + (-(0 \cdot x))) = 0,$$

so

$$0 \cdot x + 0 = 0 \cdot x = 0.$$

Similarly,

$$(x \cdot 0 + x \cdot 0) = x \cdot (0 + 0) = x \cdot 0,$$

so

$$(x \cdot 0 + x \cdot 0) + (-(0 \cdot x)) = x \cdot 0 + (-(0 \cdot x)) = 0.$$

Hence

$$x \cdot 0 + (x \cdot 0 + (-(0 \cdot x))) = 0,$$

so

$$x \cdot 0 + 0 = x \cdot 0 = 0.$$ □

This proof shows that we can get the equivalent of subtraction by adding the additive inverse. But although we can add, subtract, and multiply elements in a ring, we cannot, in general, divide elements. In fact, we can find some rings for which the product of two nonzero elements produces 0, such as $3 \cdot 2 = 0$ in the ring Z_6.

DEFINITION 9.5 If x is a nonzero element of a ring such that either $x \cdot y = 0$ or $y \cdot x = 0$ for a nonzero element y, then x is called a *zero divisor* of the ring. If a ring has no zero divisors, it is called a *ring without zero divisors.*

We see from this definition that 2 and 3 are zero divisors of the ring Z_6, since $3 \cdot 2 = 0$ in this ring. A related definition stems from the product of two elements equaling the multiplicative identity.

DEFINITION 9.6 If, for the element x in a ring with identity, there is an element y such that

$$x \cdot y = y \cdot x = e,$$

we say that x has a multiplicative inverse, or is *invertible.*

Just because an element is not a zero divisor does not mean that it is invertible. For example, 2 is not a zero divisor of the ring $\mathbb{Z}$, yet 2 is not invertible in this ring.

The smallest possible ring is the *trivial ring*, which is defined by the *Mathematica* commands

DefMultMod[**1**]
AddTable[{**0**}]
MultTable[{**0**}]

+	0
0	0

·	0
0	0

Both of these tables are displayed in GAP by the command

```
gap> MultTable([0]);

+|0
-+-
0|0
```

This ring is rather unusual because the multiplicative identity is 0. Also, 0 is actually invertible in this ring, because $0^{-1} = 0$. These two facts are true for no other ring.

DEFINITION 9.7 A ring for which every nonzero element has a multiplicative inverse is called a *division ring.*

PROPOSITION 9.5
A division ring always has a multiplicative identity and has no zero divisors.

PROOF We just saw that the trivial ring has an identity and has no zero divisors, so we may assume that the ring has a nonzero element y. Then y has a multiplicative inverse z, so we have $y \cdot z = e$, the identity. Thus, every division ring must have an identity.

Now suppose that $x \cdot y = 0$ in a division ring, with both x and y nonzero. Then y has a multiplicative inverse z, so that $y \cdot z = e$. But then

$$x = x \cdot e = x \cdot (y \cdot z) = (x \cdot y) \cdot z = 0 \cdot z = 0,$$

which contradicts the fact that x is nonzero. Thus, a division ring has no zero divisors. □

DEFINITION 9.8 A nontrivial division ring for which $x \cdot y = y \cdot x$ for all x and y is called a *field.* A division ring for which multiplication is not commutative is called a *skew field.*

We can now classify each possible type of ring. For example, the ring $\mathbb{Z}$ is a commutative ring with an identity and without zero divisors. The ring of even integers, however, has no identity element, so we would call this a commutative ring without zero divisors. Both $\mathbb{Q}$ and $\mathbb{R}$ satisfied all 13 properties, so these two rings are fields. The ring Z_6 has zero divisors, so we would call this a commutative ring with identity. The quaternions $\mathbb{H}$ have all the properties of a field except that multiplication is not commutative, so this is an example of a skew field.

9.3 Entering Finite Rings into GAP and *Mathematica*

In the first eight chapters, we entered finite groups into *Mathematica* by using the generators of the group. If we consider a finite ring simply as an abelian group under addition, we can find a set of generators B for this group (ignoring the multiplicative structure). For each element in B we determine the additive order of the element. That is, for each generator x we want to find the smallest number n such that

$$\underbrace{x + x + \cdots + x + x}_{n \text{ times}} = 0.$$

DEFINITION 9.9 If n is a positive integer, and x is any element in a ring, we define nx inductively by letting $1x = x$, and

$$nx = (n-1)x + x.$$

We also define $(-n)x$ to be $-(nx)$ for n a positive integer. Finally, we define $0x = 0$.

Because "multiplication by an integer" is merely a shorthand for repeated addition, we immediately see that

$$(m+n)x = mx + nx \qquad \text{and} \qquad (mn)x = m(nx)$$

for any element x and any integers n and m.

LEMMA 9.2
Let x and y be any two elements in a ring, and let n be an integer. Then

$$(nx) \cdot y = n(x \cdot y) = x \cdot (ny).$$

PROOF We will proceed by induction. The statement is certainly true for $n = 0$ or $n = 1$. Suppose that the statement is true for the previous case $n - 1$. But then

$$((n-1)x) \cdot y + x \cdot y = (n-1)(x \cdot y) + x \cdot y = x \cdot ((n-1)y) + x \cdot y.$$

Hence, by the distributive law,

$$((n-1)x + x) \cdot y = ((n-1)+1)(x \cdot y) = x \cdot ((n-1)y + y),$$

and so

$$(nx) \cdot y = n(x \cdot y) = x \cdot (ny).$$

Hence, the statement is true for all positive integers.

For negative integers, we can merely show that

$$(nx) \cdot y + ((-n)x) \cdot y = (nx + (-n)x) \cdot y = ((n-n)x) \cdot y = 0 \cdot y = 0.$$

$$n(x \cdot y) + (-n)(x \cdot y) = (n-n)(x \cdot y) = 0(x \cdot y) = 0.$$

$$x \cdot (ny) + x \cdot ((-n)y) = x \cdot (ny + (-n)y) = x \cdot ((n-n)y) = x \cdot 0 = 0.$$

Thus,$((-n)x) \cdot y$, $(-n)(x \cdot y)$, and $x \cdot ((-n)y)$ are the additive inverses of $(nx) \cdot y$, $n(x \cdot y)$, and $x \cdot (ny)$, respectively. But since these latter three are equal for positive n, we have

$$((-n)x) \cdot y = (-n)(x \cdot y) = x \cdot ((-n)y).$$

Hence the lemma is proven for all integers n. □

We can now use this notation within *Mathematica* to generate a finite ring. To define a ring whose additive group is isomorphic to

$$Z_{15}^* = \{1, 2, 4, 7, 8, 11, 13, 14\},$$

we find two elements that generate this group: $a = 2$ and $b = 14$.

We see that $a^4 = 1$ and $b^2 = 1$ in this group. But in our new notation, we write $4a = 0$ and $2b = 0$, since 0 is the additive identity of the ring.

To define this group in *Mathematica*, we define both $4a$ and $2b$ to be 0. The following three commands do this.

InitRing
Define[4 a, 0]
Define[2 b, 0]

TABLE 9.3: Addition table for the ring **R**

$+$	0	a	$2a$	$3a$	b	$a+b$	$2a+b$	$3a+b$
0	0	a	$2a$	$3a$	b	$a+b$	$2a+b$	$3a+b$
a	a	$2a$	$3a$	0	$a+b$	$2a+b$	$3a+b$	b
$2a$	$2a$	$3a$	0	a	$2a+b$	$3a+b$	b	$a+b$
$3a$	$3a$	0	a	$2a$	$3a+b$	b	$a+b$	$2a+b$
b	b	$a+b$	$2a+b$	$3a+b$	0	a	$2a$	$3a$
$a+b$	$a+b$	$2a+b$	$3a+b$	b	a	$2a$	$3a$	0
$2a+b$	$2a+b$	$3a+b$	b	$a+b$	$2a$	$3a$	0	a
$3a+b$	$3a+b$	b	$a+b$	$2a+b$	$3a$	0	a	$2a$

This defines the group structure of the ring. The eight elements of the group are denoted as follows:

R = AddGroup[{a,b}]
$\{0, a, 2a, 3a, b, a+b, 2a+b, 3a+b\}$

The addition table can be displayed using **AddTable[R]**, producing table 9.3.

The first statement, **InitRing**, tells *Mathematica* that we are defining a ring instead of a group. This allows the use of the plus sign instead of the dot for the additive operation. The additive identity will always be 0, so this is not needed in the command.

Although this defines the additive group very quickly, we must be selective in choosing the generators. Suppose we had instead chosen the generators $a = 2$ and $b = 7$. These two elements generate the group Z_{15}^*, but both are of order 4. So the *Mathematica* commands for entering these two generators would be

InitRing
Define[4 a, 0]
Define[4 b, 0]
R = AddGroup[{a, b}]
$\{0, a, 2a, 3a, b, a+b, 2a+b, 3a+b, 2b, a+2b, 2a+2b, 3a+2b, 3b, a+3b,$
$2a+3b, 3a+3b\}$

This gives 16 elements instead of 8! The problem is that *Mathematica* is not using the identity $2a = 2b$, which is true since $2^2 = 7^2$ (Mod 15). One solution would be to add an additional *Mathematica* command defining $2a = 2b$, but this produces some potential problems later on. A better solution is simply to make the following restriction on the set of generators.

DEFINITION 9.10 Let G be an abelian group. A *basis* is a set $B = \{x_1, x_2, x_3, \ldots x_k\}$ which generates the group such that the only way in which

$$n_1x_1 + n_2x_2 + n_3x_3 + \cdots + n_kx_k = 0$$

for integers $n_1, n_2, n_3, \ldots n_k$ is if

$$n_1x_1 = n_2x_2 = n_3x_3 = \cdots = n_kx_k = 0.$$

For a finite group, it is clear that every combination of the form

$$n_1x_1 + n_2x_2 + n_3x_3 + \cdots + n_kx_k,$$

where each n_i is non-negative and less then the order of x_i, forms a distinct element. Also, every element of G could be put in that form. Thus, the product of the orders of all the elements of B equals the order of the group.

It should be noted that *any* finite abelian group has a basis, as shown in problem 9.30.

Once we have found a basis for the additive group, and have defined the additive structure into *Mathematica*, we are ready to consider the multiplicative definitions. If we have two generators $\{a, b\}$, we will need to define $2^2 = 4$ multiplications: $a \cdot a$, $a \cdot b$, $b \cdot a$, and $b \cdot b$. These four products could be defined to be any of the elements of the ring. Thus, for ring with the additive structure of Z_{15}^*, there are up to $8^4 = 4096$ ways to finish defining the ring! However, very few of these ways of defining the products will satisfy both the distributive laws and the associative laws. Here is an example of a set of definitions that does not produce such a contradiction:

```
InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, b]
Define[a.b, 0]; Define[b.a, 0]
R = Ring[{a, b}]
```

The addition table was given above in table 9.3, while the multiplication table is given by

```
MultTable[R]
```

producing table 9.4.

The tedious task of verifying the distributive and associative laws can be handled by *Mathematica* by the command

```
CheckRing[{a, b}]
```

Notice that it suffices to give *Mathematica* just the basis for the additive group. This allows **CheckRing** to run much faster than if the entire ring were used for the argument.

TABLE 9.4: Multiplication table for the ring $\mathbf{R}$

$\cdot$	0	a	$2a$	$3a$	b	$a+b$	$2a+b$	$3a+b$
0	0	0	0	0	0	0	0	0
a	0	a	$2a$	$3a$	0	a	$2a$	$3a$
$2a$	0	$2a$	0	$2a$	0	$2a$	0	$2a$
$3a$	0	$3a$	$2a$	a	0	$3a$	$2a$	a
b	0	0	0	0	b	b	b	b
$a+b$	0	a	$2a$	$3a$	b	$a+b$	$2a+b$	$3a+b$
$2a+b$	0	$2a$	0	$2a$	b	$2a+b$	b	$2a+b$
$3a+b$	0	$3a$	$2a$	a	b	$3a+b$	$2a+b$	$a+b$

To enter a ring into GAP, we can use the `InitRing` command to identify the names of the generators, as the `FreeGroup` command did for groups. Then we define the ring in one step using the `DefineRing` command, which takes three arguments: the name of the new ring, a list showing the orders of the generators, and an array defining the possible products of two of the generators. For example, if `"a"` and `"b"` are the two generators, then the array would consist of `[[a*a, a*b],[b*a, b*b]]`. To define the ring that we defined in *Mathematica*, we would enter

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,0],[0,b]]);
gap> List(R);
[ 0*a, b, a, a+b, 2*a, 2*a+b, 3*a, 3*a+b ]
gap> CheckRing(R);
This is a ring.
gap> AddTable(R);

+     |0*a     b       a       a+b     2*a     2*a+b   3*a     3*a+b
------+-----------------------------------------------------------
0*a   |0*a     b       a       a+b     2*a     2*a+b   3*a     3*a+b
b     |b       0*a     a+b     a       2*a+b   2*a     3*a+b   3*b
a     |a       a+b     2*a     2*a+b   3*a     3*a+b   0*a     b
a+b   |a+b     a       2*a+b   2*a     3*a+b   3*a     b       0*a
2*a   |2*a     2*a+b   3*a     3*a+b   0*a     b       a       a+b
2*a+b |2*a+b   2*a     3*a+b   3*a     b       0*a     a+b     a
3*a   |3*a     3*a+b   0*a     b       a       a+b     2*a     2*a+b
3*a+b |3*a+b   3*a     b       0*a     a+b     a       2*a+b   2*a

gap> MultTable(R);

*     |0*a     b       a       a+b     2*a     2*a+b   3*a     3*a+b
------+-----------------------------------------------------------
0*a   |0*a     0*a     0*a     0*a     0*a     0*a     0*a     0*a
b     |0*a     b       0*a     b       0*a     b       0*a     b
a     |0*a     0*a     a       a       2*a     2*a     3*a     3*a
a+b   |0*a     b       a       a+b     2*a     2*a+b   3*a     3*a+b
2*a   |0*a     0*a     2*a     2*a     0*a     0*a     2*a     2*a
2*a+b |0*a     b       2*a     2*a+b   0*a     b       2*a     2*a+b
3*a   |0*a     0*a     3*a     3*a     2*a     2*a     a       a
3*a+b |0*a     b       3*a     3*a+b   2*a     2*a+b   a       a+b
```

We notice several things from this example. First of all, the zero element is listed as `0*a`, not just 0. GAP interprets 0 to mean only the integer 0, so the zero element of a ring needs a different notation. Of course, $0 \cdot a$ would give us the zero element for any generator a, so GAP picks the first generator mentioned.

As with *Mathematica*, the command `CheckRing` will see whether the object constructed obeys the distributive and associative laws. The command

```
gap> Identity(R);
a+b
```

will search the ring for a multiplicative identity. There is such an identity in this ring, even though we did not use the identity element to construct the ring. The corresponding *Mathematica* command is

FindIdent[**R**]

The multiplication table shows that many elements of R do not have inverses. Hence, this is not a division ring. Nonetheless, GAP can try to take inverses of some of the elements.

```
gap> (3*a+b)^-1;
3*a+b
gap> (2*a+b)^-1;
fail
```

We can try to define a non-commutative ring using Z_{15}^* as the additive group. If $a \cdot b = b$, yet $b \cdot a = 2a$, then the ring will not be commutative. To define this in *Mathematica*, we type in the following:

InitRing
Define[**4 a, 0**]
Define[**2 b, 0**]
Define[**a.b, b**]
Define[**b.a, 2 a**]
Define[**a.a, ???**]
Define[**b.b, ???**]
CheckRing[{**a, b**}]

or in GAP by

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[???, b],[2*a, ???]]);
gap> CheckRing(R);
```

There are actually two ways of replacing the `???`'s with elements so that a ring is formed. Here are several attempts to fill in the `???`'s.

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[0, b],[2*a, 0]]);
gap> CheckRing(R);
Associative law does not hold.
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a+b, b],[2*a, a]]);
gap> CheckRing(R);
Ring is not left distributive.
```

It would seem as though there would be 64 possibilities to check, but we can narrow the search by using the associative property. For example, $(a \cdot b) \cdot a$ must be $a \cdot (b \cdot a)$, so $2a = 2a^2$. This forces a^2 to be either a or $3a$. With this extra information, try to see if you can fill in the ??? using either GAP or *Mathematica* so that a ring is formed.

It is relatively easy to see why such a ring cannot have an identity element. GAP's `Identity` function or *Mathematica*'s **FindIdent** will return `fail`, showing that there is no identity element. In fact, every nonzero element will be a zero divisor.

PROPOSITION 9.6

If a ring with identity has an additive structure that can be generated with less than three elements, then the ring is commutative.

PROOF Suppose that x and y are two elements of the ring that generate the group under addition. That is, every element can be expressed as $mx+ny$ for integers m and n. In particular, the identity element

$$e = mx + ny$$

for some integers m and n. Since e commutes with both x and y, we have

$$mx \cdot x + ny \cdot x = (mx + ny) \cdot x = e \cdot x = x \cdot e = mx \cdot x + nx \cdot y,$$

so $ny \cdot x = nx \cdot y$.

Likewise,

$$mx \cdot y + ny \cdot y = (mx + ny) \cdot y = e \cdot y = y \cdot e = my \cdot x + ny \cdot y,$$

so $mx \cdot y = my \cdot x$.

By the greatest common divisor theorem (1.2), there are integers u and v such that

$$um + vn = \text{GCD}[m, n].$$

If we let c denote the greatest common divisor of m and n, then

$$c(x \cdot y - y \cdot x) = (um + vn)(x \cdot y - y \cdot x) = u(mx \cdot y - my \cdot x) + v(nx \cdot y - ny \cdot x) = 0.$$

What we need to show is that $(x \cdot y - y \cdot x) = 0$. The tempting thing to do is divide by c, but this operation is not allowed in rings. Instead, we will again

utilize the identity element. Since $c = \text{GCD}[m, n]$ there are integers a and b such that $m = ac$ and $n = bc$. Then

$$\begin{aligned} x \cdot y - y \cdot x = e \cdot (x \cdot y - y \cdot x) &= (acx + bcy) \cdot (x \cdot y - y \cdot x) \\ &= (ax + by) \cdot (c(x \cdot y - y \cdot x)) = (ax + by) \cdot 0 = 0. \end{aligned}$$

So $x \cdot y = y \cdot x$, and the ring is commutative. □

If we were to find a non-commutative ring with an identity, we need an additive group that requires more than two generators to define. The smallest such group is Z_{24}^*. We may suppose that the additive group is generated by the multiplicative identity e, along with two other elements a and b. Suppose that $a \cdot b = a$, while $b \cdot a = b$. This would make the ring non-commutative. We still need to discern what a^2 and b^2 should be. But $a^2 = (a \cdot b) \cdot a = a \cdot (b \cdot a) = a \cdot b = a$, and $b^2 = (b \cdot a) \cdot b = b \cdot (a \cdot b) = b \cdot a = b$.

The *Mathematica* command for defining this ring would be

```
InitRing
Define[2 e, 0]
Define[2 a, 0]
Define[2 b, 0]
Define[e.e, e]
Define[e.a, a]
Define[e.b, b]
Define[a.e, b]
Define[b.e, b]
Define[a.b, a]
Define[b.a, b]; Define[a.a, a]; Define[b.b, b]
CheckRing[{a, b}]
```

Likewise, the GAP commands would be

```
gap> InitRing("e","a","b");
gap> DefineRing("R",[2,2,2],[[e,a,b],[a,a,a],[b,b,b]]);
gap> CheckRing(R);
This is a ring.
gap> Identity{R};
e
gap> MultTable(R);
```

*	|0*e	b	a	a+b	e	e+b	e+a	e+a+b
0*e	|0*e	0*e	0*e	0*e	0*e	0*e	0*e	0*e
b	|0*e	b	b	0*e	b	0*e	0*e	b
a	|0*e	a	a	0*e	a	0*e	0*e	a
a+b	|0*e	a+b	a+b	0*e	a+b	0*e	0*e	a+b
e	|0*e	b	a	a+b	e	e+b	e+a	e+a+b
e+b	|0*e	0*e	a+b	a+b	e+b	e+b	e+a	e+a
e+a	|0*e	a+b	0*e	a+b	e+a	e+b	e+a	e+b
e+a+b	|0*e	a	b	a+b	e+a+b	e+b	e+a	e

9.4 Some Properties of Rings

One of the simplest rings to study are the rings Z_n for $n > 1$. We have already learned how to define the addition structure in *Mathematica* with a **DefSumMod** command, and the multiplication can be defined using a **DefMultMod** command. We actually can define both of these at once in *Mathematica* with the command

DefMod[15]

This defines both the addition and multiplication operations at the same time. The elements of Z_{15} are

Z15 = Ring[{1}]
$\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}$

since the group is generated by the identity element, 1. We can perform simple operations in Z_{15} such as

7 + 9
7 . 9
1/7

The GAP commands that perform these calculations are

```
gap> (7+9) mod 15;
1
gap> (7*9) mod 15;
3
gap> 1/7 mod 15;
13
```

This last operation shows that we can take multiplicative inverses of some of the elements. Even though multiplicative inverses are not guaranteed to exist for rings, some elements may be invertible.

LEMMA 9.3
Let x be an element in a ring with identity. Then if x has a multiplicative inverse, the inverse is unique. We denote the multiplicative inverse of x by x^{-1}.

PROOF Suppose that y and z are two inverses of x. Then

$$y = y \cdot e = y \cdot (x \cdot z) = (y \cdot x) \cdot z = e \cdot z = z,$$

which is a contradiction. □

PROPOSITION 9.7
If R has an identity, then the invertible elements of R form a group under multiplication. This group is denoted R^*.

PROOF Since the identity element is invertible, R^* is non-empty. Also, if x is invertible, then $(x^{-1})^{-1} = x$, so x^{-1} is also in R^*. Finally, if x and y are both invertible, then since

$$(x \cdot y) \cdot (y^{-1} \cdot x^{-1}) = x \cdot x^{-1} = e,$$

we see that $x \cdot y$ is invertible. Thus, the set of invertible elements forms a group. □

From this, we can find out when Z_n is in fact a field. The first step is to determine when Z_n will have zero divisors.

PROPOSITION 9.8
For $n > 1$, the ring Z_n has no zero divisors if, and only if, n is prime.

PROOF First suppose that n is not prime. Then we can express $n = ab$, where a and b are less then n. If e represents the identity element of Z_n, we would then have

$$(ae) \cdot (be) = (ab)(e \cdot e) = (ab)e = ne = 0.$$

But since a and b are both less than n, (ae) and (be) are both nonzero. Hence, these would both be zero divisors in Z_n.

Now suppose that n is prime, and that there are two nonzero elements (ae) and (be) such that $(ae) \cdot (be) = 0$. Then

$$(ae) \cdot (be) = (ab)(e \cdot e) = (ab)e = 0.$$

This would imply that ab is a multiple of n. But since n is prime, we would have to conclude that either a or b is a multiple of n. But this contradicts the fact that both (ae) and (be) are nonzero. Thus, if n is prime, there are no zero divisors in Z_n. □

Even if n is not prime, one of the observations that can be made while studying Z_n is that the zero divisors were precisely the nonzero elements that did not have an inverse. This is true for many of the rings we have studied.

LEMMA 9.4
Let a, b, and c be elements of a ring. If a is nonzero, and is not a zero divisor, and

$$a \cdot b = a \cdot c,$$

then $b = c$. Likewise, if

$$b \cdot a = c \cdot a$$

for a nonzero and not a zero divisor, then $b = c$. This is called the cancellation law for multiplication.

PROOF The tempting thing to do is to multiply both sides of the equation by a^{-1}. But the inverse of a may not exist, so we have to use the properties of rings instead.

If $a \cdot b = a \cdot c$ then we have

$$0 = a \cdot b - a \cdot c = a \cdot (b - c).$$

But since a is not a zero-divisor and is nonzero, we must have that $b - c = 0$. Hence $b = c$.

Likewise, if $b \cdot a = c \cdot a$, then

$$0 = b \cdot a - c \cdot a = (b - c) \cdot a$$

and since a is nonzero and not a zero divisor, $b - c = 0$, and so $b = c$. □

Notice that in the ring $\mathbb{Z}$, the element 2 is not invertible, but neither is it a zero divisor. This example seems to break the pattern that we have been observing, but also notice that $\mathbb{Z}$ is an *infinite* ring. Perhaps if we consider only *finite* rings we will be able to prove a relationship between zero divisors and invertible elements.

PROPOSITION 9.9

Let R be a finite ring. If b is a nonzero element of R which is not a zero divisor, then R has an identity element and b has a multiplicative inverse in R. Hence, every nonzero element in R is either a zero divisor or is invertible.

PROOF To utilize the fact that R is finite, let us construct a sequence of powers of b:

$$\{b^1, b^2, b^3, \ldots\}.$$

Since R is finite, two elements of this sequence must be equal, say $b^m = b^n$ for $m < n$. Using the law of cancellation, we have $b^{m-1} = b^{n-1}$. Continuing this way, we eventually get $b = b^{n-m+1}$. (It is tempting to use lemma 9.4 one more time to get $e = b^{n-m}$, but unfortunately we have yet to prove that R has an identity.)

If we now let $a = n - m + 1$, we have that $a > 1$ and $b^a = b$.

Next, let us show that b^{a-1} is an identity element in R. For any element x in R, we have

$$x \cdot b^a = x \cdot b,$$

and since b is nonzero and not a zero divisor, we can use the law of cancellation to get

$$x \cdot b^{a-1} = x.$$

Likewise, since $b^a \cdot x = b \cdot x$, we have that $b^{a-1} \cdot x = x$. Hence, there is an identity element in R, namely b^{a-1}.

Finally, we need to construct an inverse for the element b. If $a = 2$, then we have just shown that $b = e$, and hence b is its own inverse. If $a > 2$, consider the element b^{a-2}. We have that

$$b^{a-2} \cdot b = b^{a-1} = e \qquad \text{and} \qquad b \cdot b^{a-2} = b^{a-1} = e.$$

So b^{a-2} is the multiplicative inverse of b. □

COROLLARY 9.1

Every finite ring without zero divisors is a division ring.

PROOF The trivial ring is already considered to be a division ring, so we may assume that the ring is nontrivial. Then there exists a nonzero element that is not a zero divisor, so by proposition 9.9, the ring has an identity. Also by proposition 9.9, every nonzero element will have a multiplicative inverse, so the ring is a division ring. □

We finally can determine which Z_n are fields.

COROLLARY 9.2

The ring Z_n is a field if, and only if, n is prime.

PROOF If $n = 1$, then the ring $Z_n = Z_1$ is the trivial ring, which we did not consider to be a field. We may suppose that $n > 1$. If n is prime, then by proposition 9.8 Z_n has no zero divisors, and so by corollary 9.1 Z_n is a division ring. Since Z_n is obviously commutative, this tells us that Z_n is a field.

Now suppose that $n > 1$ and n is not prime. By proposition 9.8, Z_n has zero divisors, which cannot exist in a field according to proposition 9.5. Therefore Z_n is a field if, and only if, n is prime. □

To conclude this chapter, let us find an example of each of the 11 different types of rings that could exist. First we define the two rings T_4 in table 9.5 and T_8 in table 9.6. Then every ring will fall into one of the categories given in table 9.7.

TABLE 9.5: The non-commutative ring T_4

$+$	0	a	b	c
0	0	a	b	c
a	a	0	c	b
b	b	c	0	a
c	c	b	a	0

$\cdot$	0	a	b	c
0	0	0	0	0
a	0	a	a	0
b	0	b	b	0
c	0	c	c	0

TABLE 9.6: The smallest non-commutative ring T_8 with an identity

$+$	0	e	a	b	c	d	f	g
0	0	e	a	b	c	d	f	g
e	e	0	d	f	g	a	b	c
a	a	d	0	c	b	e	g	f
b	b	f	c	0	a	g	e	d
c	c	g	b	a	0	f	d	e
d	d	a	e	g	f	0	c	b
f	f	b	g	e	d	c	0	a
g	g	c	f	d	e	b	a	0

$\cdot$	0	e	a	b	c	d	f	g
0	0	0	0	0	0	0	0	0
e	0	e	a	b	c	d	f	g
a	0	a	a	a	0	0	0	a
b	0	b	b	b	0	0	0	b
c	0	c	c	c	0	0	0	c
d	0	d	0	c	c	d	f	f
f	0	f	c	0	c	d	f	d
g	0	g	b	a	c	d	f	e

TABLE 9.7: Examples for each possible type of ring

Type	Name	Example(s)
I	The trivial ring	Only one such ring, $\{0\}$.
II	Fields	$\mathbb{R}$, $\mathbb{Q}$, Z_p with p prime.
III	Skew fields	$\mathbb{H}$ = the quaternions.
IV	Commutative rings w/ identity and w/o zero divisors, but are not fields	$\mathbb{Z}$, polynomials. These rings are called *integral domains*.
V	Non-commutative rings w/ identity and w/o zero divisors, but are not skew fields	Integer quaternions: $a + bI + cJ + dK$, with $a, b, c, d \in \mathbb{Z}$.
VI	Commutative rings w/o identity and w/o zero divisors	Even integers, multiples of n, $n > 1$.
VII	Non-commutative rings w/o identity and w/o zero divisors	Even Quaternions.
VIII	Commutative rings w/ identity and w/ zero divisors	Z_n whenever $n > 1$ and n is not prime.
IX	Non-commutative rings w/ identity and w/ zero divisors	T_8 in table 9.6.
X	Commutative rings w/o identity and w/ zero divisors	The subset $\{0, 2, 4, 6\}$ of Z_8.
XI	Non-commutative rings w/o identity and w/ zero divisors	T_4 in table 9.5.

Problems for Chapter 9

Interactive Problems

9.1 Notice that in *Mathematica*, the plot of rational numbers between 0.03 and 0.1,

Z = ShowRationals[0.03, 0.1]

shows most of the points lying on a curve. Try to find the equation of this curve, using the fact that each dot is three fourths closer to the x-axis than the previous dot. Verify your answer by plotting the curve with the points, using the following command:

Show[Z, Plot[*function goes in here* , {x, 0.03, 0.1}]]

Hint: Scale the function so that $f(0.1) = 1$.

9.2 Use *Mathematica* or GAP to define a ring of order 2 that has no identity element. Show both the addition table and the multiplication table.

9.3 Use *Mathematica* or GAP to find a non-commutative ring of order 8, for which the additive group is isomorphic to Z_{24}^*, formed from the basis $\{a, b, c\}$, and for which $a \cdot b = a$, $b \cdot a = b$, $a \cdot c = c$, and $c \cdot a = a$.

Hint: Using the associative law, determine what a^2, b^2, and c^2 must be. Then show that $c \cdot b$ must commute with a. Use trial and error to determine $b \cdot c$.

9.4 Define in GAP or *Mathematica* the smallest non-commutative ring, T_4 defined by table 9.5.

9.5 Define in GAP or *Mathematica* the smallest non-commutative ring with an identity, T_8 defined by table 9.6.

Hint: The basis can be chosen to be e, a, and b.

Non-Interactive Problems

9.6 Prove that the square root of 3 is irrational.

9.7 Prove that the cube root of 2 is irrational.

9.8 Prove that if a is rational and b is irrational, then $a + b$ is irrational.

9.9 Prove that between any two distinct real numbers, there is an irrational number.

Hint: Use problem 9.8 along with proposition 9.1.

9.10 Prove that if a is rational and nonzero, and b is irrational, then $a \cdot b$ is irrational.

9.11 Prove that $y = \sqrt{2} + \sqrt{3}$ is irrational.
Hint: First show that y^2 is irrational.

9.12 Is the sum of two irrational numbers always irrational? If not, find a counter-example.

9.13 For the quaternions, $\mathbb{H}$, we define the *conjugate* of an element $x = a + bi + cj + dk$ to be $\overline{x} = a - bi - cj - dk$. Prove that $\overline{x_1} + \overline{x_2} = \overline{x_1 + x_2}$ for all x_1 and x_2 in $\mathbb{H}$.

9.14 Prove or disprove: $\overline{x_1} \cdot \overline{x_2} = \overline{x_1 \cdot x_2}$ for all x_1 and x_2 in $\mathbb{H}$. (See problem 9.13.)

9.15 Prove that for x in $\mathbb{H}$, $x \cdot \overline{x} = \overline{x} \cdot x = a^2 + b^2 + c^2 + d^2$. (See problem 9.13.)

9.16 For all x in $\mathbb{H}$, we define the *absolute value* of x to be $|x| = \sqrt{x \cdot \overline{x}}$. Prove that $|x_1 \cdot x_2| = |x_1|\,|x_2|$. (See problem 9.13.)

9.17 Prove or disprove: For all x in the quaternions $\mathbb{H}$, $(x+1)\cdot(x-1) = x^2 - 1$.

9.18 Prove or disprove: For all x in the quaternions $\mathbb{H}$, $(x+i)\cdot(x-i) = x^2 + 1$.

9.19 Let

$$\mathbb{Z}[\sqrt{2}] = \{x + y\sqrt{2} \quad | \quad x, y \in \mathbb{Z}\}.$$

Prove that $\mathbb{Z}[\sqrt{2}]$ is a ring under the ordinary addition and multiplication of real numbers.

9.20 Prove that a ring can have at most one multiplicative identity.

9.21 Show that the non-commutative ring T_4 given by table 9.5 has two elements r such that $x \cdot r = x$ for all x in the ring, yet has no element for which $r \cdot x = x$ for all x in the ring.

9.22 Prove that a ring with a cyclic additive group must be commutative.

9.23 Prove that if n is an integer, and x is an element of a ring, then $n(-x) = -(nx)$.

9.24 Let x be an element of a commutative ring R which has an inverse x^{-1}. Let y be another element of R such that $y^2 = 0$. Prove that $x + y$ has an inverse in R.

9.25 Suppose that G is an abelian group with respect to addition. Define a multiplication on G by $x \cdot y = 0$ for all x and y in G. Show that G forms a ring.

9.26 Find a specific example of two elements x and y in a ring R such that $x \cdot y = 0$, but $y \cdot x$ is nonzero.

Hint: Which of the 11 types of rings would R have to be?

9.27 Let R be a ring for which $x^2 = x$ for all x in the ring. Prove that $-x = x$ for all elements x. Such rings are called *Boolean* rings.

9.28 Let R be a ring for which $x^2 = x$ for all x in the ring. Prove that the ring R is commutative. (See problem 9.27.)

9.29 Define new operations of addition and multiplication in $\mathbb{Z}$ by $x \oplus y = x + y - 1$ and $x \otimes y = x + y - xy$. Verify that $\mathbb{Z}$ forms a ring with respect to these new operations.

9.30 Use the fundamental theorem of abelian groups (6.2) to show that every finite abelian group has a basis.

9.31 An element a in a ring R is *idempotent* if $a^2 = a$. Prove that a nontrivial division ring must contain exactly two idempotent elements.

9.32 Show that if R is a commutative ring, and x and y are elements of R, then

$$(x + y)^2 = x^2 + 2xy + y^2$$

and

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

9.33 Let R be a commutative ring. Define the *binomial coefficient*

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot (n-2) \cdots (n-k+1)}{1 \cdot 2 \cdot 3 \cdots k}, (0 \leq k \leq n).$$

Using induction, prove the *binomial theorem* in R:

$$(x + y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \cdots + \binom{n}{n}y^n.$$

This page intentionally left blank

Chapter 10

The Structure within Rings

10.1 Subrings

It is natural to ask whether we can have smaller rings within a larger ring, just as we saw smaller groups inside of a larger group. This suggests the following definition.

DEFINITION 10.1 Let R be a ring. A non-empty subset S is a *subring* if S is a ring with respect to the addition $(+)$ and multiplication $(\cdot)$ of R.

We have already seen some examples of subrings. For example, the set of even integers is a ring contained in the ring of integers, which is contained in the ring of rational numbers, which in turn is contained in the ring of real numbers. The next proposition gives us a quick way to determine if a subset is indeed a subring.

PROPOSITION 10.1
A non-empty subset S is a subring of a ring R if, and only if, whenever x and y are in S, $x - y$ and $x \cdot y$ are in S.

PROOF Certainly if S is a subring, then $x - y$ and $x \cdot y$ would be in S whenever x and y are in S. So let us suppose that S is non-empty, and is closed with respect to subtraction and multiplication. If x is any element in S, then $x - x = 0$ is in S, so S contains an additive identity. Also, $0 - x = -x$ would also be in S, so S contains additive inverses of all of its elements. Then whenever x and y are in S, $x - (-y) = x + y$ is in S, so S is closed with respect to addition. The commutative and associative properties of addition, as well as the associative and two distributive laws for multiplication, come from the original ring R. Finally, S is closed with respect to multiplication, so S is a subring. □

Notice that from the definition every nontrivial ring R will contain at least two subrings: the trivial ring $\{0\}$ will be a subring, as well as the entire ring

TABLE 10.1: Tables for the subring S

$+$	0	a	$2a$	$3a$
0	0	a	$2a$	$3a$
a	a	$2a$	$3a$	0
$2a$	$2a$	$3a$	0	a
$3a$	$3a$	0	a	$2a$

$\cdot$	0	a	$2a$	$3a$
0	0	0	0	0
a	0	a	$2a$	$3a$
$2a$	0	$2a$	0	$2a$
$3a$	0	$3a$	$2a$	a

R. These two subrings are called the *trivial subrings.*

Let us look at an example. Here is the ring of order 8 we defined by tables 9.3 and 9.4:

InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, b]
Define[a.b, 0]; Define[b.a, 0]
R = Ring[{a, b}]

The set

S = {0, a, 2a, 3a}

can be seen to be a subring from the addition and multiplication tables in table 10.1. To generate these tables in GAP, we use the following commands:

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,0],[0,b]]);
gap> S := Ring(a);
<ring with 1 generators>
gap> List(S);
[ 0*a, a, 2*a, 3*a ]
gap> AddTable(S);

+  |0*a  a    2*a  3*a
---+-----------------
0*a|0*a  a    2*a  3*a
a  |a    2*a  3*a  0*a
2*a|2*a  3*a  0*a  a
3*a|3*a  0*a  a    2*a

gap> MultTable(S);

*  |0*a  a    2*a  3*a
---+-----------------
0*a|0*a  0*a  0*a  0*a
a  |0*a  a    2*a  3*a
2*a|0*a  2*a  0*a  2*a
3*a|0*a  3*a  2*a  a
```

One can see that S is closed with respect to both addition and multiplication. Furthermore, additive inverses exist for all elements, so S is also closed with respect to subtraction. Thus, by proposition 10.1, this is a subring.

Ironically, the subring S has an identity element,

FindIdent[S]

```
gap> Identity(S);
a
```

which is different than the identity element for R. In general the existence of a multiplicative identity of a subring is totally independent of the multiplicative identity of R.

Recall that the intersection of a number of subgroups was again a subgroup. We could ask whether the same is true for subrings.

PROPOSITION 10.2

Given any non-empty collection of subrings of the group R, denoted by L, then the intersection of all of the subrings in the collection

$$H^* = \bigcap_{H \in L} H$$

is a subring of R.

PROOF First of all, note that H^* is not the empty set, since 0 is in each H in the collection. We now can apply proposition 10.1. Let x and y be two elements in H^*. Then, for every $H \in L$, we have $x, y \in H$.

Since each H is a subring of R, we have $x - y \in H$ and $x \cdot y \in H$ for all $H \in L$. Therefore, $x - y$ and $x \cdot y$ are in H^*, and so H^* is a subring of R. □

As with subgroups, we now have a general method of producing subrings of a ring R. Let S be any subset of R. We can consider the collection L of all subrings of R that contain the set P. This collection is non-empty since it contains the subring R itself. So by proposition 10.2,

$$[S] = H^* = \bigcap_{H \in L} H$$

is a subring of R. By the way that the collection was defined, $[S]$ contains S. Actually, $[S]$ is the *smallest* subring of R containing the subset S.

DEFINITION 10.2 We call $[S]$ the subring of R *generated* by the set S.

Just as in the case for the `Group` command, the command `Ring` finds $[S]$ for any set S in either *Mathematica*® or GAP. For example, we can find some subrings for the non-commutative group of order 8,

InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, 0]
Define[a.b, b]; Define[b.a, 2 a]
R = Ring[{a, b}]

with the commands

Ring[{0}]
Ring[{a}]
Ring[{2a}]
Ring[{2a, b}]

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,b],[2*a,0]]);
gap> List(Ring(0*a));
[ 0*a ]
gap> List(Ring(a));
[ 0*a, a, 2*a, 3*a ]
gap> List(Ring(2*a));
[ 0*a, 2*a ]
gap> List(Ring(2*a,b));
[0*a, b, 2*a, 2*a+b]
```

In this way, we can find all subrings of the ring R. In fact, GAP has a command `Subrings` that finds all of the possible subrings.

```
gap> L := Subrings(R);
[ <ring with 1 generators>, <ring with 1 generators>,
  <ring with 1 generators>, <ring with 1 generators>,
  <ring with 2 generators>, <ring with 2 generators>,
  <ring with 2 generators>, <ring with 3 generators> ]
gap> List(L, List);
[ [ 0*a ], [ 0*a, b ], [ 0*a, 2*a ], [ 0*a, 2*a+b ],
  [ 0*a, b, 2*a, 2*a+b ], [ 0*a, a, 2*a, 3*a ],
  [ 0*a, a+b, 2*a, 3*a+b ],
  [ 0*a, b, a, a+b, 2*a, 2*a+b, 3*a, 3*a+b ] ]
```

So we see that there are six nontrivial subrings for this ring.

We can easily find all of the subrings for the infinite ring $\mathbb{Z}$.

PROPOSITION 10.3

A subring of the ring of integers $\mathbb{Z}$ consists of all multiples of some nonnegative number n. This subring is denoted $n\mathbb{Z}$.

PROOF First of all, the trivial subring $\{0\}$ can be considered the set of all multiples of 0. Also, the entire ring $\mathbb{Z}$ could be considered all of the multiples of 1. Let S be a nontrivial subring, and let x be in S. Then $-x$ is also in S, so S must contain some positive integers. Let n be the smallest positive integer

contained in S. Certainly all multiples of n would be in S, but suppose that some element m in S is not a multiple of n. Then by the greatest common divisor theorem (1.2), there exist two integers u and v such that

$$un + vm = \text{GCD}(n, m)$$

Since S is closed under addition, this implies that $\text{GCD}(n, m)$ is in S. But m is not a multiple of n, so $\text{GCD}(n, m) < n$. But this contradicts the fact that n is the *smallest* positive integer in S. Thus, S consists exactly of all of the multiples of n, and so $S = n\mathbb{Z}$. □

Although the subrings of $\mathbb{Z}$ are easily classified, this is not the case with the ring of real numbers. Consider the set S of all numbers of the form

$$x + y\sqrt{2}$$

where x and y are rational numbers. We can have *Mathematica* verify that the product of two such numbers

```
ClearDefs
Expand[(x1 + y1 2^(1/2)) (x2 + y2 2^(1/2) )]
```

produces a number in this form. Since S is obviously closed with respect to subtraction, S is a subring of R.

To define this subring in GAP, we can let e represent 1, and a represent $\sqrt{2}$. These two elements are both of infinite additive order. We can convey this to GAP by entering "0" for the order of each of the elements. Then $a^2 = 2e$, so the ring can be entered into GAP by the commands

```
gap> InitRing("e","a");
gap> DefineRing("R",[0,0],[[e,a],[a,2*e]]);
gap> Size(R);
infinity
gap> (e+2*a)*(4*e-3*a);
-8*e+5*a
```

This last statement demonstrates that

$$(1 + 2\sqrt{2}) \cdot (4 - 3\sqrt{2}) = -8 + 5\sqrt{2}.$$

Clearly, the subrings of the real numbers can be much more complicated than the subrings of the integers.

10.2 Quotient Rings and Ideals

When we studied group theory, one of the most important concepts we discovered was being able to form a quotient group out of the cosets of certain

subgroups—namely the normal subgroups. A natural question is whether it is possible to form quotient rings out of the cosets of a subring.

Let us look at an example. Here is the non-commutative ring of order 8 from the last section.

InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, 0]
Define[a.b, b]; Define[b.a, 2 a]
R = Ring[{a, b}]

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,b],[2*a,0]]);
```

We found this ring has six nontrivial subrings.

$$S_1 = \{0, a, 2a, 3a\}, \qquad S_2 = \{0, 2a\}, \qquad S_3 = \{0, b\},$$
$$S_4 = \{0, a+b, 2a, 3a+b\}, \quad S_5 = \{0, 2a+b\}, \quad S_6 = \{0, 2a, b, 2a+b\}.$$

We would expect the additive structure of the quotient ring to be the additive quotient group R/S. We can use *Mathematica* or GAP to find the cosets of S under the operation of addition. Since left and right cosets are the same when working with rings, we will simply use the `Coset` command for both GAP and *Mathematica*.

S1 = {0, a, 2a, 3a}
Q = Coset[R, S1]

```
gap> S1 := Ring(a);
<ring with 1 generators>
gap> Q := Coset(R,S1);
[ [ 0*a, a, 2*a, 3*a ], [ b, a+b, 2*a+b, 3*a+b ] ]
```

We can *add* two cosets together using the following definition:

$$X + Y = \{x + y \quad | \quad x \in X \quad \text{and} \quad y \in Y\}.$$

This gives us a natural way to add the elements of the quotient Q, which is shown in table 10.2.

TABLE 10.2: Addition for the quotient ring Q

$+$	$\{0, a, 2a, 3a\}$	$\{b, a+b, 2a+b, 3a+b\}$
$\{0, a, 2a, 3a\}$	$\{0, a, 2a, 3a\}$	$\{b, a+b, 2a+b, 3a+b\}$
$\{b, a+b, 2a+b, 3a+b\}$	$\{b, a+b, 2a+b, 3a+b\}$	$\{0, a, 2a, 3a\}$

This table can be produced using the **AddTable[Q]** command in either *Mathematica* or GAP.

```
gap> AddTable(Q);
+                  |[0*a,a,2*a,3*a]       [b,a+b,2*a+b,2*a+b]
-------------------+-----------------------------------------
[0*a,a,2*a,3*a]    |[0*a,a,2*a,3*a]       [b,a+b,2*a+b,2*a+b]
[b,a+b,2*a+b,2*a+b]|[b,a+b,2*a+b,2*a+b]   [0*a,a,2*a,3*a]
```

The natural way to define the product of two sets is the way we defined such a product for groups:

$$X \cdot Y = \{x \cdot y \quad | \quad x \in X \quad \text{and} \quad y \in Y\}.$$

Will such a product of two cosets in Q yield another coset? Here is GAP's response:

```
gap> MultTable(Q);
*                  |[0*a,a,2*a,3*a]       [b,a+b,2*a+b,2*a+b]
-------------------+-----------------------------------------
[0*a,a,2*a,3*a]    |[0*a,a,2*a,3*a]
[b,a+b,2*a+b,2*a+b]|[0*a,a,2*a,3*a]
```

Unfortunately no! The multiplication tables in *Mathematica* or GAP reveal black or blank squares—which indicate that the product of two cosets is not a coset. The problem lies in the product

{0, a, 2a, 3a} . {b, a+b, 2a+b, 3a+b}

```
gap> Mult(R,[0*a,a,2*a,3*a],[b,a+b,2*a+b,3*a+b]);
[ 0*a, b, a+b, 2*a, 2*a+b, 3*a+b ]
```

which produces extra elements. To ensure that S acts as the zero element in the product of cosets, we need to have S times any element of R needs to produce only elements in S.

Suppose we found a subring S for which $S \cdot x$ always was a subset of S. By the same argument we would also require that $x \cdot S$ be a subset of S. Using *Mathematica* or GAP

S2 = {0,2a}
S2 . R
R . S2

```
gap> S2 := Ring(2*a);
<ring with 1 generators>
gap> Mult(R,S2,R);
[ 0*a, 2*a ]
gap> Mult(R,R,S2);
[ 0*a, 2*a ]
```

we see that both $R \cdot S_2$ and $S_2 \cdot R$ are subsets of S_2, so this ensures that the additive identity of the quotient group $\{0, 2a\}$ will behave as the zero element in the product of cosets. The multiplication table for the quotient group is as given by the commands

Q = Coset[R, S2]
MultTable[Q]

which produce table 10.3.

TABLE 10.3: Multiplying cosets of S_2

$\cdot$	$\{0, 2a\}$	$\{a, 3a\}$	$\{b, 2a+b\}$	$\{a+b, 3a+b\}$
$\{0, 2a\}$	$\{0\}$	$\{0, 2a\}$	$\{0\}$	$\{0, 2a\}$
$\{a, 3a\}$	$\{0, 2a\}$	$\{a, 3a\}$	$\{b, 2a+b\}$	$\{a+b, 3a+b\}$
$\{b, 2a+b\}$	$\{0\}$	$\{0, 2a\}$	$\{0\}$	$\{0, 2a\}$
$\{a+b, 3a+b\}$	$\{0, 2a\}$	$\{a, 3a\}$	$\{b, 2a+b\}$	$\{a+b, 3a+b\}$

The corresponding GAP commands are

```
gap> Q := Coset(R,S2);
[ [ 0*a, 2*a ], [ b, 2*a+b ], [ a, 3*a ], [ a+b, 3*a+b ] ]
gap> MultTable(Q);
```

*	[0*a,2*a]	[b,2*a+b]	[a,3*a]	[a+b,3*a+b]
[0*a,2*a]	[0*a]	[0*a]	[0*a,2*a]	[0*a,2*a]
[b,2*a+b]	[0*a]	[0*a]	[0*a,2*a]	[0*a,2*a]
[a,3*a]	[0*a,2*a]	[b,2*a+b]	[a,3*a]	[a+b,3*a+b]
[a+b,3*a+b]	[0*a,2*a]	[b,2*a+b]	[a,3*a]	[a+b,3*a+b]

This multiplication table is non-commutative, even though all of the subrings of R are commutative. So this quotient is unlike any of the subrings of R.

However, not every product yields a coset—sometimes it yields only a *subset* of a coset. One way to rectify this slight blemish in our multiplication table is to add the identity coset to each entry in the table. That is, instead of defining the product of the cosets X and Y to be $X \cdot Y$, we define the product of two cosets to be

$$X * Y = X \cdot Y + S.$$

The command

QuotientRing = True

creates a multiplication table using this new definition of the product of two cosets. Thus, **MultTable[Q]** produces a similar table as table 10.3, only every $\{0\}$ is replaced by $\{0, 2a\}$.

```
gap> QuotientRing := true;
true
gap> MultTable(Q);
```

*	[0*a,2*a]	[b,2*a+b]	[a,3*a]	[a+b,3*a+b]
[0*a,2*a]	[0*a,2*a]	[0*a,2*a]	[0*a,2*a]	[0*a,2*a]
[b,2*a+b]	[0*a,2*a]	[0*a,2*a]	[0*a,2*a]	[0*a,2*a]
[a,3*a]	[0*a,2*a]	[b,2*a+b]	[a,3*a]	[a+b,3*a+b]
[a+b,3*a+b]	[0*a,2*a]	[b,2*a+b]	[a,3*a]	[a+b,3*a+b]

The key to getting the quotient ring to work lies in the fact that $S_2 \cdot R$ and $R \cdot S_2$ were subsets of S_2. Let us first define the special type of subring that will allow quotient rings.

DEFINITION 10.3 A subring I of a ring R is called an *ideal* of R if both $I \cdot R$ and $R \cdot I$ are contained in the subring I.

We already observed that if a subring is not an ideal, then the quotient ring cannot be defined. Let us now show that a quotient ring can be defined provided that I is an ideal.

PROPOSITION 10.4
Let R be a ring, and let I be an ideal of R. Then the additive quotient group R/I forms a ring, with the product of two cosets X and Y being $X * Y = X \cdot Y + I$. This ring is called the quotient ring R/I.

PROOF The quotient group R/I is an abelian group, so we need only to check that the multiplication is closed, and that the associativity and two distributive laws hold.

Let X and Y be two cosets of R/I. Let x be an element in X, and y an element in Y. Then the product of the cosets X and Y is

$$X * Y = X \cdot Y + I = (x + I) \cdot (y + I) + I = x \cdot y + I \cdot y + x \cdot I + I \cdot I + I.$$

Because I is an ideal, $I \cdot y$, $x \cdot I$, and $I \cdot I$ are all subsets of I. Hence, the sum $I \cdot y + x \cdot I + I \cdot I + I$ will be a subset of I. But since the last term of this expression is I, $I \cdot y + x \cdot I + I \cdot I + I$ contains the ideal I, so this sum equals I. Thus,

$$(x + I) * (y + I) = X * Y = X \cdot Y + I = x \cdot y + I,$$

which is a coset of R/I.

Now suppose that X, Y, and Z are three cosets of R/I with x, y, and z being representative elements, respectively. Then

$$\begin{aligned}
(X * Y) * Z &= ((x + I) * (y + I)) * (z + I) \\
&= (x \cdot y + I) * (z + I) \\
&= ((x \cdot y) \cdot z + I) \\
&= (x \cdot (y \cdot z) + I) \\
&= (x + I) * (y \cdot z + I) \\
&= (x + I) * ((y + I) * (z + I)) \\
&= X * (Y * Z).
\end{aligned}$$

So multiplication is associative. Also,

$$\begin{aligned} X * (Y + Z) &= (x + I) * (y + z + I) \\ &= (x(y + z) + I) \\ &= x \cdot y + x \cdot z + I \\ &= (x \cdot y + I) + (x \cdot z + I) \\ &= X * Y + X * Z, \end{aligned}$$

and

$$\begin{aligned} (X + Y) * Z &= (x + y + I) * (z + I) \\ &= ((x + y) \cdot z + I) \\ &= x \cdot z + y \cdot z + I \\ &= (x \cdot z + I) + (y \cdot z + I) \\ &= X * Z + Y * Z. \end{aligned}$$

Thus, the two distributive laws hold, so R/I is a ring. □

This shows that the ideals play the same role for rings that normal subgroups did for groups, namely that subsets with an additional property allow for quotients to be defined.

Let us consider the ideals of the ring $\mathbb{Z}$. By proposition 10.3, all subrings are of the form $S = n\mathbb{Z}$ for some n. Yet any multiple of n times an integer yields a multiple of n, so $S \cdot \mathbb{Z} = \mathbb{Z} \cdot S = S$. Therefore, every subring of $\mathbb{Z}$ is an ideal.

The cosets of the quotient ring $\mathbb{Z}/(n\mathbb{Z})$ can be expressed in the form

$$a + n\mathbb{Z},$$

where $a = 0, 1, 2, \ldots n - 1$. Clearly the quotient ring behaves exactly like the ring Z_n. We say that the quotient ring is *isomorphic* to Z_n.

In contrast, let us consider a ring like the rational numbers $\mathbb{Q}$. Even though there are a host of subrings of $\mathbb{Q}$, the only ideals are the trivial subrings. This can be generalized by the following proposition.

PROPOSITION 10.5

Any field or skew field can only have trivial ideals.

PROOF Let K be a field or skew field, and suppose that there is a nontrivial ideal I of K. Then there is a nonzero element x in I, and hence x^{-1} exists in K. Thus

$$1 = x \cdot x^{-1} \in I \cdot K \subseteq I.$$

So the multiplicative identity 1 is contained in I. But then,

$$K = 1 \cdot K \subseteq I \cdot K \subseteq I.$$

Hence, $I = K$, so the only ideals of K are the trivial ideals. □

We have already observed that the intersection of two subrings is again a subring. The natural question is whether the intersection of two ideals gives an ideal.

PROPOSITION 10.6

If L is a non-empty collection of ideals of a ring R, then the intersection of all of these ideals

$$I^* = \bigcap_{I \in L} I$$

is an ideal of R.

PROOF Since I^* is an intersection of subrings of R, by proposition 10.2 I^* is a subring of R. Thus, we only need to check that $I^* \cdot R$ and $R \cdot I^*$ are contained in I^*.

Suppose that x is an element of I^*. Then x is in each $I \in L$, and so $x \cdot R$ and $R \cdot x$ are subsets of each I in the collection. Thus, $x \cdot R$ and $R \cdot x$ will both be subsets of I^*. Since this result is true for every x in I^*, we have that $I^* \cdot R$ and $R \cdot I^*$ are both subsets of I^*. Therefore, I^* is an ideal. □

We can now define the smallest ideal of R that contains a subset S. We proceed as we did for subrings, and consider the collection L of all ideals of R containing S. Then the smallest ideal of R containing S would be

$$(S) = \bigcap_{I \in L} I.$$

We call (S) the *ideal generated by S*. Notice the distinction between this notation and the notation $[S]$ of the subring generated by S. If S contains only one element, say a, we will use the notation (a) rather than the cumbersome $(\{a\})$ to denote the ideal generated by a.

This proposition allows us to quickly find all ideals of a ring. For example, in the non-commutative ring R of order 8, which we were working with above in this section, we can have *Mathematica* or GAP find (S) using the command

Ideal[R, S]

for different subsets S. For example, when $S = \{a\}$,

```
gap> I := Ideal(R,[a]);
<two-sided ideal in <ring with 2 generators>, (1 generators)>
gap> List(I);
[ 0*a, b, a, a+b, 2*a, 2*a+b, 3*a, 3*a+b ]
```

we find that this command produces the whole ring, so a cannot be contained in any nontrivial ideal. Likewise, $3a$, $a+b$, and $3a+b$ cannot be in a nontrivial ideal. The three remaining nonzero elements, $2a$, b, and $2a + b$, generate different ideals.

```
gap> List(Ideal(R,[2*a]));
[ 0*a, 2*a ]
gap> List(Ideal(R,[b]));
[ 0*a, b, 2*a, 2*a+b ]
gap> List(Ideal(R,[2*a+b]));
[ 0*a, 2*a+b ]
```

These three ideals will be denoted by $(2a)$, (b), and $(2a + b)$. It is clear that any ideal containing two out of three of these elements must contain b, and therefore must be (b). Hence, there are exactly five ideals in this ring: the two trivial ideals that can be denoted (0) and (a), and the three ideals $(2a)$, (b), and $(2a + b)$. We can verify this in GAP with the command `Ideals`, which gives a list of all the ideals of a finite ring.

```
gap> L := Ideals(R);
[ <ring with 1 generators>, <ring with 1 generators>,
  <ring with 1 generators>, <ring with 2 generators>,
  <ring with 3 generators> ]
gap> List(L, List);
[ [ 0*a ], [ 0*a, 2*a ], [ 0*a, 2*a+b ], [ 0*a, b, 2*a, 2*a+b ],
  [ 0*a, b, a, a+b, 2*a, 2*a+b, 3*a, 3*a+b ] ]
```

Notice that all five ideals can be generated with only one element.

DEFINITION 10.4 An ideal of R that is generated by only one element of R is called a *principal ideal*. If all of the ideals of R are principal ideals, then the ring is called a *principal ideal ring*.

The ring of integers $\mathbb{Z}$ is a principal ideal ring, since all ideals (in fact all subrings) are of the form $n\mathbb{Z}$, which is generated by the single element n. Since $\mathbb{Z}$ is also an integral domain, we will combine the two terms and call $\mathbb{Z}$ a *principal ideal domain*, or *PID*. We will talk more about PIDs in section 12.3.

10.3 Ring Isomorphisms

As we work with different rings, it is natural to ask whether we can consider two rings to be "equivalent" if the elements of one ring can be renamed to form the other ring. We have already seen that the quotient ring $\mathbb{Z}/(n\mathbb{Z})$ was essentially the same ring as Z_n. We will proceed the same way we defined isomorphisms with groups.

DEFINITION 10.5 Let A and B be two rings. A *ring isomorphism* from A to B is a one-to-one mapping $f : A \to B$ such that

$$f(x + y) = f(x) + f(y) \qquad \text{and}$$
$$f(x \cdot y) = f(x) \cdot f(y)$$

for all $x, y, \in A$. If there exists a ring isomorphism from A to B that is surjective, then we say that the rings A and B are *isomorphic,* denoted by $A \approx B$.

For example, we can define a mapping for the quotient ring $\mathbb{Z}/(n\mathbb{Z})$ as follows:

$$f(a + n\mathbb{Z}) = a \ (\text{Mod } n), \qquad a = 0, 1, 2, \ldots, (n - 1).$$

Then clearly f is an injective and surjective function from $\mathbb{Z}/(n\mathbb{Z})$ to Z_n. Furthermore, $f(a + b) = f(a) + f(b)$, and $f(a \cdot b) = f(a) \cdot f(b)$. So we have that $\mathbb{Z}/(n\mathbb{Z}) \approx Z_n$.

Let us look at another example of a ring isomorphism. Consider the following *Mathematica* commands that define a ring of order 10, and produce the addition and multiplication tables shown in table 10.4.

```
InitRing
Define[10 a, 0]
Define[a.a, 2 a]
A = Ring[{a}]
AddTable[A]
MultTable[A]
```

TABLE 10.4: Addition and multiplication in the ring A

$+$	0	a	$2a$	$3a$	$4a$	$5a$	$6a$	$7a$	$8a$	$9a$
0	0	a	$2a$	$3a$	$4a$	$5a$	$6a$	$7a$	$8a$	$9a$
a	a	$2a$	$3a$	$4a$	$5a$	$6a$	$7a$	$8a$	$9a$	0
$2a$	$2a$	$3a$	$4a$	$5a$	$6a$	$7a$	$8a$	$9a$	0	a
$3a$	$3a$	$4a$	$5a$	$6a$	$7a$	$8a$	$9a$	0	a	$2a$
$4a$	$4a$	$5a$	$6a$	$7a$	$8a$	$9a$	0	a	$2a$	$3a$
$5a$	$5a$	$6a$	$7a$	$8a$	$9a$	0	a	$2a$	$3a$	$4a$
$6a$	$6a$	$7a$	$8a$	$9a$	0	a	$2a$	$3a$	$4a$	$5a$
$7a$	$7a$	$8a$	$9a$	0	a	$2a$	$3a$	$4a$	$5a$	$6a$
$8a$	$8a$	$9a$	0	a	$2a$	$3a$	$4a$	$5a$	$6a$	$7a$
$9a$	$9a$	0	a	$2a$	$3a$	$4a$	$5a$	$6a$	$7a$	$8a$

$\cdot$	0	a	$2a$	$3a$	$4a$	$5a$	$6a$	$7a$	$8a$	$9a$
0	0	0	0	0	0	0	0	0	0	0
a	0	$2a$	$4a$	$6a$	$8a$	0	$2a$	$4a$	$6a$	$8a$
$2a$	0	$4a$	$8a$	$2a$	$6a$	0	$4a$	$8a$	$2a$	$6a$
$3a$	0	$6a$	$2a$	$8a$	$4a$	0	$6a$	$2a$	$8a$	$4a$
$4a$	0	$8a$	$6a$	$4a$	$2a$	0	$8a$	$6a$	$4a$	$2a$
$5a$	0	0	0	0	0	0	0	0	0	0
$6a$	0	$2a$	$4a$	$6a$	$8a$	0	$2a$	$4a$	$6a$	$8a$
$7a$	0	$4a$	$8a$	$2a$	$6a$	0	$4a$	$8a$	$2a$	$6a$
$8a$	0	$6a$	$2a$	$8a$	$4a$	0	$6a$	$2a$	$8a$	$4a$
$9a$	0	$8a$	$6a$	$4a$	$2a$	0	$8a$	$6a$	$4a$	$2a$

The multiplicative structure of this group is different than Z_{10}, since there is no multiplicative identity. Yet the additive group is isomorphic to the group Z_{10}. This is not surprising, since there is only one abelian group of order 10.

We can easily find other rings of order 10. Suppose we let b be the generator of the additive group, and define $b^2 = 6b$.

Define[10 b, 0]
Define[b.b, 6 b]
B = Ring[{b}]

The addition table is virtually the same as for the ring A, but the multiplication table looks different. This time let us load both rings into GAP, using different generators for the two rings.

```
gap> InitRing("a");
gap> DefineRing("A",[10],[[2*a]]);
gap> InitRing("b");
gap> DefineRing("B",[10],[[6*b]]);
gap> List(A);
[ 0*a, a, 2*a, 3*a, 4*a, 5*a, 6*a, 7*a, 8*a, 9*a ]
gap> List(B);
[ 0*b, b, 2*b, 3*b, 4*b, 5*b, 6*b, 7*b, 8*b, 9*b ]
gap> MultTable(B);

*  |0*b  b    2*b  3*b  4*b  5*b  6*b  7*b  8*b  9*b
---+----------------------------------------------
0*b|0*b  0*b  0*b  0*b  0*b  0*b  0*b  0*b  0*b  0*b
b  |0*b  6*b  2*b  8*b  4*b  0*b  6*b  2*b  8*b  4*b
2*b|0*b  2*b  4*b  6*b  8*b  0*b  2*b  4*b  6*b  8*b
3*b|0*b  8*b  6*b  4*b  2*b  0*b  8*b  6*b  4*b  2*b
4*b|0*b  4*b  8*b  2*b  6*b  0*b  4*b  8*b  2*b  6*b
5*b|0*b  0*b  0*b  0*b  0*b  0*b  0*b  0*b  0*b  0*b
6*b|0*b  6*b  2*b  8*b  4*b  0*b  6*b  2*b  8*b  4*b
7*b|0*b  2*b  4*b  6*b  8*b  0*b  2*b  4*b  6*b  8*b
8*b|0*b  8*b  6*b  4*b  2*b  0*b  8*b  6*b  4*b  2*b
9*b|0*b  4*b  8*b  2*b  6*b  0*b  4*b  8*b  2*b  6*b
```

There are enough similarities between A and B to ask whether they are isomorphic. It is not immediately clear what the isomorphism should be. One way to help find an isomorphism between A and B is to show that both of these are isomorphic to a subring of the Z_n for some n. For example, consider $2Z_{20}$, the even elements of Z_{20}.

```
gap> InitRing("e");
gap> DefineRing("Z20",[20],[[e]]);
gap> R := Ring(2*e);
<ring with 1 generators>
gap> List(R);
[ 0*e, 2*e, 4*e, 6*e, 8*e, 10*e, 12*e, 14*e, 16*e, 18*e ]
```

or in *Mathematica*,

TABLE 10.5: Multiplication in $2Z_{20}$

$\cdot$	0	$2e$	$4e$	$6e$	$8e$	$10e$	$12e$	$14e$	$16e$	$18e$
0	0	0	0	0	0	0	0	0	0	0
$2e$	0	$4e$	$8e$	$12e$	$16e$	0	$4e$	$8e$	$12e$	$16e$
$4e$	0	$8e$	$16e$	$4e$	$12e$	0	$8e$	$16e$	$4e$	$12e$
$6e$	0	$12e$	$4e$	$16e$	$8e$	0	$12e$	$4e$	$16e$	$8e$
$8e$	0	$16e$	$12e$	$8e$	$4e$	0	$16e$	$12e$	$8e$	$4e$
$10e$	0	0	0	0	0	0	0	0	0	0
$12e$	0	$4e$	$8e$	$12e$	$16e$	0	$4e$	$8e$	$12e$	$16e$
$14e$	0	$8e$	$16e$	$4e$	$12e$	0	$8e$	$16e$	$4e$	$12e$
$16e$	0	$12e$	$4e$	$16e$	$8e$	0	$12e$	$4e$	$16e$	$8e$
$18e$	0	$16e$	$12e$	$8e$	$4e$	0	$16e$	$12e$	$8e$	$4e$

Define[20 e, 0]
Define[e.e, e]
R = Ring[{2 e}]
MultTable[R]

which produces table 10.5. (The reason why we did not use the **DefMod** command to load Z_{20} in *Mathematica* is because we would erase the rings A and B.) In *Mathematica* one can see that the color patterns for A and R are the same, so that $A \approx 2Z_{20}$. To prove this in GAP, we can construct a function f mapping A to Z_{20} using the `RingHomomorphismByImages` command, which works basically the same as its group counterpart. Since A has only one generator, a, we tell GAP where a will be sent to, which is obviously $2e$.

```
gap> f := RingHomomorphismByImages(A,R,[a],[2*e]);
[ a ] -> [ 2*e ]
gap> List(Image(f));
[ 0*e, 2*e, 4*e, 6*e, 8*e, 10*e, 12*e, 14*e, 16*e, 18*e ]
```

As one might expect after working with group homomorphisms, if we can prove that a function is a homomorphism, and that it is one-to-one, then we have an isomorphism. Since GAP successfully defined a homomorphism, and the image contains 10 elements, then it must be one-to-one, so GAP has verified that $A \approx 2Z_{20}$. We can now generalize this example as follows.

PROPOSITION 10.7

Let R be a finite ring whose additive structure is a cyclic group of order n. Let x be a generator of the additive group. Then $x^2 = k \cdot x$ for some positive integer $k \leq n$, and

$$A \approx kZ_{kn}.$$

PROOF If $x^2 = 0$, we can let $k = n$, so that k will be positive and $k \cdot x = 0 = x^2$. If x^2 is not zero, then since x generates the additive group, there is a k such that $x^2 = k \cdot x$ with $0 < k < n$.

Now the natural mapping is one that sends $f(a \cdot x) = k \cdot a \text{ (Mod } kn)$. This is obviously one-to-one and onto, since the value of a ranges from 0 to $n - 1$. To check that this is an isomorphism, note that

$$\begin{aligned} f(a \cdot x + b \cdot x) &= f((a + b) \cdot x) = k \cdot (a + b) \text{ (Mod } kn) \\ &= k \cdot a \text{ (Mod } kn) + k \cdot b \text{ (Mod } kn) \\ &= f(a \cdot x) + f(b \cdot x). \end{aligned}$$

Also,

$$\begin{aligned} f((a \cdot x) \cdot (b \cdot x)) &= f(a \cdot b \cdot x^2) \\ &= f(a \cdot b \cdot k \cdot x) \\ &= k \cdot a \cdot b \cdot k \text{ (Mod } kn) \\ &= (k \cdot a \text{ (Mod } kn)) \cdot (k \cdot b \text{ (Mod } kn)) \\ &= f(a \cdot x) \cdot f(b \cdot x). \end{aligned}$$

Therefore, f is an isomorphism, and $R \approx kZ_{kn}$. □

This proposition shows not only that $A \approx 2Z_{20}$, but also that $B \approx 6A_{60}$, since $b^2 = 6b$ in this ring.

DEFINITION 10.6 A *cyclic ring* is a ring whose additive group is cyclic.

Note that this definition of cyclic rings also includes the infinite rings $\mathbb{Z}$ and its subrings $k\mathbb{Z}$.

In order to prove that in fact $A \approx B$, we will need a few lemmas about number theory. Once these are proven, we will be able to determine *all* non-isomorphic rings of order 10.

LEMMA 10.1

Let d be a positive divisor of n, and let f be the largest divisor of d that is coprime to (n/d). Then if q is coprime to both f and (n/d), then q is coprime to n.

PROOF Suppose that $\text{GCD}(q, n)$ is not 1. Then there is a prime number p that divides neither f nor (n/d), yet divides n. Thus, p must divide d.

Now $f \cdot p$ will be coprime to (n/d) since both f and p are. Also, since f is not a multiple of p while d is, $f \cdot p$ will be a divisor of d. But we defined f to be the *largest* factor of d coprime to (n/d). This contradiction shows that $\text{GCD}(q, n) = 1$. □

LEMMA 10.2
Given two positive numbers x and y, there exist u and v in $\mathbb{Z}$ such that

$$ux + vy = \text{GCD}(x, y),$$

where u is coprime to y.

PROOF The greatest common divisor theorem (1.2) would give us values for u and v, but there would be no way to guarantee that u would be coprime to y.

Let $k = \text{GCD}(x, y)$. Then (x/k) and (y/k) are coprime, so (x/k) has an multiplicative inverse in $Z_{(y/k)}$, say n. That is,

$$\frac{x}{k} \cdot n \equiv 1 \left(\text{Mod } \frac{y}{k}\right).$$

Let f be the largest divisor of k that is coprime to (y/k). By the Chinese remainder theorem (1.3), there is a number u such that

$$u \equiv n \left(\text{Mod } \frac{y}{k}\right)$$

and

$$u \equiv 1 \text{ (Mod } f).$$

Since n is coprime to (y/k), u is coprime to (y/k). Also, u is coprime to f, so by lemma 10.1 u is coprime to y. Also,

$$u \cdot \frac{x}{k} \equiv 1 \left(\text{Mod } \frac{y}{k}\right)$$

so there is a v such that $u \cdot \frac{x}{k} + v \cdot \frac{y}{k} = 1$. Multiplying both sides by k gives us

$$u \cdot x + v \cdot y = k = \text{GCD}(x, y). \qquad \square$$

THEOREM 10.1: The Cyclic Ring Theorem
If x and n are positive integers, then

$$xZ_{x \cdot n} \approx kZ_{k \cdot n},$$

where $k = \text{GCD}(x, n)$.

PROOF Since $k = \text{GCD}(x, n)$ by lemma 10.2 we can find integers u and v such that $u \cdot x + v \cdot n = k$, where u is coprime to n. We now define a mapping f from kZ_{kn} to xZ_{xn} as follows:

$$f(k \cdot w \text{ (Mod } kn)) = u \cdot x \cdot w \text{ (Mod } xn).$$

Note that this is well defined, since if $k \cdot w$ is equivalent to $k \cdot p \pmod{kn}$ then

$$\begin{aligned} w \equiv p \pmod{n} &\Longrightarrow x \cdot w \equiv x \cdot p \pmod{xn} \\ &\Longrightarrow u \cdot x \cdot w \equiv u \cdot x \cdot p \pmod{xn}. \end{aligned}$$

Next we need to show that f is a homomorphism from kZ_{kn} to xZ_{xn}. If $a = k \cdot w \pmod{kn}$ and $b = k \cdot z \pmod{kn}$, then

$$\begin{aligned} f(a+b) &= f(k \cdot w + k \cdot z \pmod{kn}) = u \cdot (x \cdot w + x \cdot z) \pmod{xn} \\ &= u \cdot x \cdot w + u \cdot x \cdot z \pmod{xn} = f(a) + f(b). \end{aligned}$$

$$\begin{aligned} f(a \cdot b) &= f(k \cdot w \cdot k \cdot z \pmod{kn}) = u \cdot x \cdot w \cdot k \cdot z \pmod{xn} \\ &= u \cdot x \cdot w \cdot (u \cdot x + v \cdot n) \cdot z \pmod{xn} \\ &= (u \cdot x \cdot w \cdot u \cdot x \cdot z + u \cdot x \cdot w \cdot v \cdot n \cdot z) \pmod{xn} \\ &= (u \cdot x \cdot w) \cdot (u \cdot x \cdot z) \pmod{xn} = f(a) \cdot f(b). \end{aligned}$$

So f is indeed a homomorphism from kZ_{kn} to xZ_{xn}.

Since u is coprime to n, u has an inverse, $u^{-1} \pmod{n}$. Then we see that f is onto, since any element $x \cdot a \pmod{xn}$ in xZ_{xn} can be obtained by taking

$$f(k \cdot a \cdot u^{-1} \pmod{kn}) = u \cdot x \cdot a \cdot u^{-1} \pmod{xn} = x \cdot a \pmod{xn}.$$

Finally, both xZ_{xn} and kZ_{kn} contain n elements, so by the pigeonhole principle f must be a one-to-one function. Thus, f is an isomorphism, and $xZ_{xn} \approx kZ_{kn}$. □

Because $2 = \text{GCD}(6, 10)$, we see that $A \approx 2Z_{20}$ is isomorphic to $B \approx 6Z_{60}$. But what is the isomorphism? Theorem 10.1 does not explicitly give a formula for where a should map to in B, so we have to use trial and error. Since a is an additive generator of A, we know that it should map to one of the additive generators of B, $\{b, 3b, 7b, 9b\}$.

```
gap> g := RingHomomorphismByImages(A,B,[a],[b]);
fail
gap> g := RingHomomorphismByImages(A,B,[a],[3*b]);
fail
gap> g := RingHomomorphismByImages(A,B,[a],[7*b]);
[ a ] -> [ 7*b ]
gap> List(Image(g));
[ 0*b, b, 2*b, 3*b, 4*b, 5*b, 6*b, 7*b, 8*b, 9*b ]
```

Since the image is all of B, GAP finally found an isomorphism between A and B.

In fact, since the only rings of order 10 are cyclic rings, there are four possible non-isomorphic rings of order 10:

$$Z_{10}, \qquad 2Z_{20}, \qquad 5Z_{50}, \qquad \text{and} \qquad 10Z_{100}.$$

It is easy to see that these rings are all distinct by looking at the multiplication tables.

COROLLARY 10.1
The number of non-isomorphic cyclic rings of order n is precisely the number of divisors of n (including 1 and n).

PROOF By proposition 10.7 every cyclic ring of order n is isomorphic to kZ_{kn} for some value of k. By the cyclic ring theorem, we see that this is isomorphic to dZ_{dn}, where $d = \text{GCD}(k, n)$. Hence d is a divisor of n. We need to show that two different rings of this form are non-isomorphic. Consider the rings $A = dZ_{dn}$ and $B = fZ_{fn}$, where d and f are different divisors of n. Perhaps the easiest way to show that these are different is to count the number of elements in A and B that can appear in the multiplication tables. The elements that can appear in the table for A are

$$d^2, 2d^2, 3d^2, \ldots, nd = 0$$

while the elements appearing in the multiplication table of B are

$$f^2, 2f^2, 3f^2, \ldots, nf = 0.$$

Thus, there are n/d such elements of A, and n/f elements of B. Since d and f are different, we see that the rings A and B are not isomorphic. Therefore, there is a one-to-one correspondence between the factors of n and the cyclic rings of order n. □

Although this corollary seems to be a big help in finding *all* finite rings, there are, in fact, many non-cyclic rings. For example, there are 8 non-cyclic rings of order 4, which when combined with the 3 cyclic rings from corollary 10.1 gives a total of 11 rings of order 4. There are 52 rings of order 8 (4 cyclic, 20 with additive group Z_{15}^*, and 28 with an additive group Z_{24}^*).

Table 10.6 shows the number of rings of a given order. There are at least 18,590 known rings of order 32, but it has not been proven that these are all of them.

In GAP, we can load any of the rings of order 8 or less. The command `NumberSmallRings` will produce the number of rings of a certain order, as given in table 10.6. Then `SmallRings` will load one of the rings. The following shows how we can load the 51st ring of order 8.

```
gap> NumberSmallRings(8);
52
gap> R := SmallRing(8,51);
<ring with 3 generators>
gap> MultTable(R);
```

```
*     |0*a     c       b       b+c     a       a+c     a+b     a+b+c
------+-------------------------------------------------------------
0*a   |0*a     0*a     0*a     0*a     0*a     0*a     0*a     0*a
c     |0*a     c       b       b+c     c       0*a     b+c     b
b     |0*a     b       b+c     c       b       0*a     c       b+c
b+c   |0*a     b+c     c       b       b+c     0*a     b       c
a     |0*a     c       b       b+c     a       a+c     a+b     a+b+c
a+c   |0*a     0*a     0*a     0*a     a+c     a+c     a+c     a+c
a+b   |0*a     b+c     c       b       a+b     a+c     a+b+c   a
a+b+c|0*a     b       b+c     c       a+b+c   a+c     a       a+b
```

TABLE 10.6: Rings of order n

n	rings	n	rings	n	rings	n	rings
1	1	9	11	17	2	25	11
2	2	10	4	18	22	26	4
3	2	11	2	19	2	27	59
4	11	12	22	20	22	28	22
5	2	13	2	21	4	29	2
6	4	14	4	22	4	30	8
7	2	15	4	23	2	31	2
8	52	16	390	24	104	32	???

10.4 Homomorphisms and Kernels

Since we defined a ring isomorphism in a similar fashion as group isomorphisms, we naturally will define ring homomorphisms by mimicking group homomorphisms.

DEFINITION 10.7 If A and B are two rings, then a mapping $f : A \to B$ such that

$$f(x + y) = f(x) + f(y),$$

and

$$f(x \cdot y) = f(x) \cdot f(y),$$

for all x and y in A is called a *ring homomorphism.*

Note that a ring homomorphism will also be a group homomorphism from the additive group of A to the additive group of B. Thus, we can immediately apply the results of group homomorphisms to see two properties of ring homomorphisms.

If f is a ring homomorphism from A to B, then

$$f(0) = 0$$

and

$$f(-x) = -f(x) \qquad \text{for all} \qquad x \in A.$$

Any isomorphism is certainly a homomorphism. But let us see how to define a homomorphism between two non-isomorphic rings. Consider a homomorphism between Z_3 and Z_6. We define Z_3 and Z_6 simultaneously by using two different generators.

InitRing
Define[3 a, 0]; Define[a.a, a]
Define[6 b, 0]; Define[b.b, b]
Z3 = Ring[{a}]
Z6 = Ring[{b}]

```
gap> InitRing("a");
gap> DefineRing("Z3",[3],[[a]]);
gap> InitRing("b");
gap> DefineRing("Z6",[6],[[b]]);
gap> List(Z3);
[ 0*a, a, 2*a ]
gap> List(Z6);
[ 0*b, b, 2*b, 3*b, 4*b, 5*b ]
```

The homomorphism is determined completely by the value of $f(a)$. A natural choice would be to let $f(a) = 2b$.

```
gap> f := RingHomomorphismByImages(Z3, Z6, [a], [2*b]);
fail
```

GAP shows that this would not produce a homomorphism. One way to correct this problem would be to send $f(a)$ to the zero element of Z_6, which GAP writes as `0*b`.

```
gap> f := RingHomomorphismByImages(Z3,Z6,[a],[0*b]);
[ a ] -> [ 0*b ]
gap> List(Image(f));
[ 0*b ]
```

or, in *Mathematica*,

Homomorph[F]
Define[F[a], 0]
CheckHomo[F, Z3]

DEFINITION 10.8 If A and B are any two rings, then the mapping $f : A \to B$

$$f(x) = 0 \qquad \text{for all } x \in A$$

is called the *zero homomorphism from A to B.*

We define $f(S)$, where S is a *set* of elements in the domain of f, to be the set of all values $f(x)$, where x is in S. We can also define the inverse image of an element y to be $f^{-1}(y)$, the set of elements such that $f(x) = y$. In fact, we can define the inverse image of a set of elements in the same way: $f^{-1}(T)$ is the set of elements such that $f(x)$ is in T. We can use *Mathematica* to find the image of a set by merely entering $F[S]$, rather than having to bother with $F[\{S\}]$ as we did with the group theory notebooks. We can find the inverse image of an element or a set in *Mathematica* just as we did for group homomorphisms.

PROPOSITION 10.8
Suppose f is a homomorphism from the ring A to the ring B. Then if S is a subring of A, then $f(S)$ is a subring of B. Likewise, if T is a subring of B, then $f^{-1}(T)$ will be a subring of A.

PROOF Suppose S is a subring of A. We will use proposition 10.1 to show that $f(S)$ is a subring of B. The element $f(0) = 0$ is in $f(S)$, so $f(S)$ is non-empty. If u and v are two elements of $f(S)$, then there exist elements x and y in S such that

$$f(x) = u$$

and

$$f(y) = v.$$

But $x \cdot y$ and $x - y$ are also in S, and so

$$f(x \cdot y) = f(x) \cdot f(y) = u \cdot v$$

and

$$f(x - y) = f(x) - f(y) = u - v$$

must be in $f(S)$. Thus, by proposition 10.1, $f(S)$ is a subring of B.

Now suppose that T is a subring of B. Since 0 is contained in $f^{-1}(T)$, we have that $f^{-1}(T)$ is non-empty. If x and y are two elements of $f^{-1}(T)$, then $f(x)$ and $f(y)$ will be two elements of T. Thus,

$$f(x \cdot y) = f(x) \cdot f(y)$$

and

$$f(x - y) = f(x) - f(y)$$

would be elements of T. Hence, $x \cdot y$ and $x - y$ are in $f^{-1}(T)$. Thus, by proposition 10.1, $f^{-1}(T)$ is a subring of A. □

We can define the kernel and the image of a homomorphism in the same way that we did for group homomorphisms.

DEFINITION 10.9 Given a homomorphism f from the ring A to the ring B, the *kernel* of f is $f^{-1}(0)$, denoted $\mathrm{Ker}(f)$. The *image* of f is $f(A)$, denoted $\mathrm{Im}(f)$.

In GAP, the kernel of a homomorphism can be found with either the `Kernel` command or the `PreImages` command.

```
gap> List(PreImages(f,0*b));
[ 0*a, a, 2*a ]
gap> List(Kernel(f));
[ 0*a, a, 2*a ]
```

In *Mathematica*, we can use the **HomoInverse** command to find the kernel of a homomorphism, or we can use the command

Kernel[F, Z3]

as we did for group homomorphisms. The images are even easier to find using *Mathematica*:

F[Z3]

When we have a homomorphism from A to B, we have by proposition 10.8 that the image will be a subring of B. Likewise, the kernel of a homomorphism will be a subring of A. However, we can say even more about the kernel.

PROPOSITION 10.9
If f is a homomorphism from the ring A to the ring B, then the kernel of f is an ideal of A. Furthermore, f is injective if, and only if, $\mathrm{Ker}(f) = \{0\}$.

PROOF Suppose that x is in the kernel of f, and y is any other element of A. Then

$$f(x \cdot y) = f(x) \cdot f(y) = 0 \cdot f(y) = 0,$$

and

$$f(y \cdot x) = f(y) \cdot f(x) = f(y) \cdot 0 = 0.$$

Hence, $x \cdot y$ and $y \cdot x$ are in the kernel of f, so the kernel is an ideal of A.

If f is injective, then $f^{-1}(0)$ can only contain one element, which must be 0. On the other hand, if $f^{-1}(0) = \{0\}$, then

$$\begin{aligned} f(x) = f(y) &\Longrightarrow f(x) - f(y) = 0 \\ &\Longrightarrow f(x - y) = 0 \\ &\Longrightarrow x - y = 0 \\ &\Longrightarrow x = y. \end{aligned}$$

Therefore, f is injective if, and only if, $\mathrm{Ker}(f) = \{0\}$. □

We have yet to find a nontrivial homomorphism from Z_3 to Z_6. Yet there is one possibility we haven't tried yet.

```
gap> g := RingHomomorphismByImages(Z3, Z6, [a], [4*b]);
[ a ] -> [ 4*b ]
gap> List(Kernel(g));
[ 0*a ]
```

This shows that Z_3 is in fact isomorphic to a subring of Z_6.

Let us look at another example of a homomorphism, considering the noncommutative ring R of order 8 used throughout section 10.2. If we wanted to define a homomorphism from R to some other ring S, the kernel would have to be an ideal of R. But R has only three nontrivial ideals:

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,b],[2*a,0]]);
gap> List(Ideals(R), List);
[ [ 0*a ], [ 0*a, 2*a ], [0*a, 2*a+b ], [ 0*a, b, 2*a, 2*a+b ],
  [ 0*a, b, a, a+b, 2*a, 2*a+b, 3*a, 3*a+b ] ]
```

InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, 0]
Define[a.b, b]; Define[b.a, 2 a]
I1 = {0, 2 a}
I2 = {0, 2 a + b}
I3 = {0, 2 a, b, 2 a + b}

To produce an interesting homomorphism, we would use one of these ideals as the kernel. To which ring should we map R?

The natural answer would be the quotient ring. Since there is a natural *group* homomorphism from R to R/I, we can ask whether this group homomorphism extends to become a ring homomorphism.

Let us define $Q = R/I_1$.

R = Ring[{a,b}];
Q = Coset[R, I1]

We wish to define a homomorphism $i(x)$ which maps an element in R to the coset of Q containing that element.

Homomorph[i]
Define[i[a], {a, 3 a}]
Define[i[b], {b, 2 a + b}]

We can see if this works with the CheckHomo command.

CheckHomo[i, R]

The kernel of this homomorphism,

Kernel[i, R]

is of course $I_1 = \{0, 2a\}$.

LEMMA 10.3
If I is an ideal of the ring R, then the natural mapping $i : R \to R/I$ defined by $i(x) = x + I$ is a surjective ring homomorphism from R to R/I with the kernel being I.

PROOF It is clear that the rule $i(x) = x + I$ defines a surjective mapping i from R to R/I, and that $\mathrm{Ker}(i) = I$. We need only to check that $i(x)$ is a homomorphism.

Since

$$\begin{aligned} i(x+y) &= (x+y)+I \\ &= (x+I)+(y+I) \\ &= i(x)+i(y) \end{aligned}$$

and

$$\begin{aligned} i(x\cdot y) &= x\cdot y+I \\ &= (x+I)\cdot(y+I) \\ &= i(x)\cdot i(y), \end{aligned}$$

we see that $i(x)$ is indeed a surjective homomorphism. □

We can define this natural homomorphism in GAP using only the ideal of the ring.

```
gap> I1 := Ring(2*a);
<ring with 1 generators>
gap> f := NaturalHomomorphismByIdeal(R,I1);
[ a, b ] -> [ q1, q2 ]
```

This actually does two things. It defines a new ring R/I, using a whole new set of generators `q1`, `q2`, ... Then it defines the map f from R to this new quotient ring. We can display the quotient ring by looking at the image of f.

```
gap> Q := Image(f);
<ring with 2 generators>
gap> MultTable(Q);

*     |0*q1   q2     q1     q1+q2
------+-------------------------
0*q1  |0*q1   0*q1   0*q1   0*q1
q2    |0*q1   0*q1   0*q1   0*q1
q1    |0*q1   q2     q1     q1+q2
q1+q2|0*q1   q2     q1     q1+q2
```

In the homomorphisms produced by lemma 10.3, the image of the homomorphism is isomorphic to $R/\text{Ker}(f)$. The first isomorphism theorem studied in the volume on groups shows that the additive group on $\text{Im}(f)$ would be group isomorphic to the additive structure of $R/\text{Ker}(f)$. It is easy to show that the ring $\text{Im}(f)$ is isomorphic to the ring $R/\text{Ker}(f)$ as well, giving us an isomorphism theorem for rings.

THEOREM 10.2: The First Ring Isomorphism Theorem
Let f be a ring homomorphism from a ring R to a ring S, whose image is H. If the kernel of f is I, then there is a natural surjective isomorphism $f : R/I \to H$ which causes the diagram in figure 10.1 to commute. (Here, $i(x)$ is the homomorphism defined in lemma 10.3.) Thus, $H \approx R/I$.

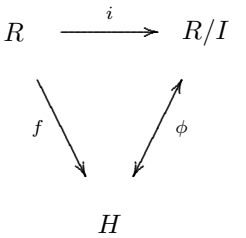


FIGURE 10.1: Commuting diagram for theorem 10.2

PROOF Figure 10.1 actually helps us determine how ϕ needs to be defined. For each coset $(x + I)$ in R/I, we need to have

$$\phi(x + I) = f(x)$$

in order for the diagram to commute. To prove that this rule defines a mapping, we need to show that this is well defined. That is, if $x + I = y + I$ it

needs to be true that $f(x) = f(y)$, or else there would be a contradiction in the definition of ϕ. But

$$\begin{aligned} x + I = y + I &\iff x - y \in I \\ &\iff f(x - y) = 0 \\ &\iff f(x) = f(y) \\ &\iff \phi(x + I) = \phi(y + I). \end{aligned}$$

So we see that the definition of ϕ will not produce any such contradictions.

To show that ϕ is a homomorphism, we have that

$$\begin{aligned} \phi((x + I) + (y + I)) &= \phi(x + y + I) \\ &= f(x + y) \\ &= f(x) + f(y) \\ &= \phi(x + I) + \phi(y + I), \end{aligned}$$

and

$$\begin{aligned} \phi((x + I) \cdot (y + I)) &= \phi(x \cdot y + I) \\ &= f(x \cdot y) \\ &= f(x) \cdot f(y) \\ &= \phi(x + I) \cdot \phi(y + I). \end{aligned}$$

So ϕ is a homomorphism from R/I to H. It is apparent that this homomorphism is onto, and

$$\begin{aligned} \phi(x + I) = 0 &\iff f(x) = 0 \\ &\iff x \in I \\ &\iff x + I = I. \end{aligned}$$

So the kernel of ϕ is $\{I\}$, the zero element of R/I. Thus, ϕ is an isomorphism from R/I onto H, so $R/I \approx H$. Since the mapping ϕ was defined so that the diagram in figure 10.1 commutes, the theorem is proved. □

It should be noted that there are second and third ring isomorphism theorems. These are considered in problems 10.46 and 10.47.

Although most of the rings we have defined in this chapter have been finite rings, it should be pointed out that whenever we defined a finite ring in *Mathematica*, we also have defined an infinite ring in the process. Consider the example of the non-commutative ring of order 8:

```
InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, 0]
Define[a.b, b]; Define[b.a, 2 a]
R = Ring[{a, b}]
```

There is no multiplicative identity for this ring. *Mathematica* can multiply any element of R by any integer, and simplify it to an element of R:

(3 a + b) . 7

But we can *add* an integer to an element of R, such as:

(3 a + b) + 7

This is not an element of R, but could this be an element of a larger ring? Suppose we let M denote the set of all expressions of the form (integer + element of R). The *Mathematica* operations cause M to be a ring in its own right.

DEFINITION 10.10 We say that the ring R is *embedded* in the ring S if there exists an injective homomorphism $f : R \to S$. The mapping f is called an *embedding* of R in S.

Mathematica has demonstrated that the ring R is embedded into a much larger ring that contains a multiplicative identity. In fact, the integers are also embedded into this ring. We can do this with any ring.

THEOREM 10.3: The Embedding Theorem
Let R be a ring. Then R can be embedded in a ring S that has a multiplicative identity.

PROOF Rather than expressing elements as an integer plus an element of R, we will use an order pair (n, x), where n is an integer and x is in R, to denote the elements of our new ring. Thus, we define S to be the set

$$S = \{(n, x) \mid n \in \mathbb{Z}, \quad x \in R\}.$$

We define addition and multiplication on S as follows:

$$(n_1, x_1) + (n_2, x_2) = (n_1 + n_2, x_1 + x_2),$$
$$(n_1, x_1) \cdot (n_2, x_2) = (n_1 n_2, x_1 \cdot x_2 + n_1 x_2 + n_2 x_1).$$

It is clear that S forms an abelian group under addition, with the zero element being $(0, 0)$. The product of two elements of S is clearly in S, so we only need to check the associativity of multiplication, and the two distributive laws. We have that

$$\begin{aligned}((n_1, x_1) \cdot (n_2, x_2)) \cdot (n_3, x_3) &= (n_1 n_2, x_1 \cdot x_2 + n_1 x_2 + n_2 x_1) \cdot (n_3, x_3) \\ &= (n_1 n_2 n_3, (x_1 \cdot x_2 + n_1 x_2 + n_2 x_1) \cdot (n_2, x_3) + \\ &\quad n_1 n_2 x_3 + n_3 (x_1 \cdot x_2 + n_1 x_2 + n_2 x_1)) \\ &= (n_1 n_2 n_3, x_1 \cdot x_2 \cdot x_3 + n_1 x_2 \cdot x_3 + n_2 x_1 \cdot x_3 + \\ &\quad n_1 n_2 x_3 + n_3 x_1 \cdot x_2 + n_1 n_3 x_2 + n_2 n_3 x_1).\end{aligned}$$

Also,

$$\begin{aligned}(n_1, x_2) \cdot ((n_2, x_2) \cdot (n_3, x_3)) &= (n_1, x_1) \cdot (n_2 n_3, x_2 \cdot x_3 + n_2 x_3 + n_3 x_2) \\ &= (n_1 n_2 n_3, x_1 \cdot (x_2 \cdot x_3 + n_2 x_3 + n_3 x_2) + \\ &\quad n_1(x_2 \cdot x_3 + n_2 x_3 + n_3 x_2) + n_2 n_3 x_1) \\ &= (n_1 n_2 n_3, x_1 \cdot n_2 x_1 \cdot x_3 + n_3 x_1 \cdot x_2 + \\ &\quad n_1 x_2 \cdot x_3 + n_1 n_2 x_3 + n_1 n_3 x_2 + n_2 n_3 x_1).\end{aligned}$$

These two are equal, so multiplication in S is associative. We also have

$$\begin{aligned}&((n_1, x_1) + (n_2, x_2)) \cdot (n_2, x_3) \\ &\quad = (n_1 + n_2, x_1 + x_2) \cdot (n_3, x_3) \\ &\quad = (n_1 n_3 + n_2 n_3, x_1 \cdot x_3 + x_2 \cdot x_3 + n_1 x_3 + n_2 x_3 + n_3 x_1 + n_3 x_2) \\ &\quad = (n_1, x_1) \cdot (n_3, x_3) + (n_2, x_2) \cdot (n_2, x_3),\end{aligned}$$

and

$$\begin{aligned}&(n_1, x_1) \cdot ((n_2, x_2) + (n_3, x_3)) \\ &\quad = (n_1, x_1) \cdot (n_2 + n_3, x_2 + x_3) \\ &\quad = (n_1 n_2 + n_1 n_3, x_1 \cdot x_2 + x_1 \cdot x_3 + n_1 x_2 + n_1 x_3 + n_2 x_1 + n_3 x_1) \\ &\quad = (n_1, x_1) \cdot (n_2, x_2) + (n_1, x_1) \cdot (n_3, x_3),\end{aligned}$$

so the two distributive laws are satisfied. Thus, S is a ring.

Furthermore, the element $(1, 0)$ in S acts as a multiplicative identity, since

$$(n, x) \cdot (1, 0) = (n \cdot 1, x \cdot 0 + n \cdot 0 + 1 \cdot x) = (n, x),$$

and

$$(1, 0) \cdot (n, x) = (1 \cdot n, 0 \cdot x + 1 \cdot x + n \cdot 0) = (n, x).$$

All that is left is to show that the ring R can be embedded into S. We can define a mapping from R to S simply by letting $f(x) = (0, x)$. This is certainly an injective mapping, and it is easy to check that

$$f(x) + f(y) = (0, x) + (0, y) = (0, x + y) = f(x, y),$$

and

$$f(x) \cdot f(y) = (0, x) \cdot (0, y) = (0, x \cdot y + 0 \cdot y + 0 \cdot x) = (0, x \cdot y) = f(x \cdot y).$$

So we have an embedding of R in S, which completes the proof. □

We call the ring S used in this theorem the *extension of R by the integers.* This ring is important because it allows us to treat any ring as though it has a multiplicative identity by using the ring S in place of the ring R.

To define the extension ring of R by the integers in GAP, we have to redefine it using an additional generator, say e, for which $e \cdot x = x \cdot e = x$ for all generators, and for which the order of e is infinite. Thus, to define the extension ring of the above example, we get

```
gap> InitRing("e","a","b");
gap> DefineRing("R",[0,4,2],[[e,a,b],[a,a,b],[b,b,2*a]]);
gap> Size(R);
infinity
```

Notice that to indicate that the generator e was of infinite order, we entered a 0 in the array position for that generator. The reason of course is that one cannot enter ∞ on the keyboard, and GAP can interpret order 0 to mean that no positive number times e will equal 0.

Problems for Chapter 10

Interactive Problems

10.1 Find all of the subrings of the ring of order 8:

InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, 0]
Define[a.b, b]; Define[b.a, 0]
R = Ring[{a, b}]

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,b],[0,0]]);
```

Which of these subrings are ideals?

10.2 Load the rings Z_{12} and Z_6 into *Mathematica* or GAP simultaneously with the commands:

InitRing
Define[12 a, 0]; Define[a.a, a]
Z12 = Ring[{a}]
Define[6 b, 0]; Define[b.b, b]
Z6 = Ring[{b}]

```
gap> InitRing("a");
gap> DefineRing("Z12",[12],[[a]]);
gap> InitRing("b");
gap> DefineRing("Z6",[6],[[b]]);
```

Show that $I = \{0, 6a\}$ is an ideal of Z_{12}, and display addition and multiplication tables of the quotient ring Z_{12}/I, showing that Z_{12}/I is isomorphic to Z_6.

10.3 Use *Mathematica* or GAP to find the eight non-isomorphic non-cyclic rings of order 4.

Hint: The additive group must be isomorphic to Z_8^*, so the ring is defined by:

InitRing
Define[2 a, 0]; Define[2 b, 0]
Define[a.a, ???]
Define[b.b, ???]
Define[a.b, ???]
Define[b.a, ???]
CheckRing[{**a, b**}]

```
gap> InitRing("a","b");
gap> DefineRing("R",[2,2],[[???,???],[???,???]]);
gap> CheckRing(R);
```

Fill in each ??? with a member of $\{0, a, b, a + b\}$ to see whether a ring is formed. Is there a faster way than trying all $4^4 = 256$ combinations?

10.4 Use *Mathematica* or GAP to display the multiplication tables of all rings of order 6.

Non-Interactive Problems

10.5 Let y be an element of a ring R. Let

$$A = \{x \in R \quad | \quad x \cdot y = 0\}.$$

Show that A is a subring of R.

10.6 Show that $2\mathbb{Z} \cup 3\mathbb{Z}$ is not a subring of $\mathbb{Z}$. (The symbol $\cup$ denotes the *union* of the two sets.)

10.7 If X and Y are ideals of a ring, show that the *sum* of X and Y,

$$X + Y = \{x + y \quad | \quad x \in X \text{ and } y \in Y\}$$

is an ideal.

10.8 In the ring of integers, find a positive integer n such that

$$(n) = (12) + (16).$$

(See problem 10.7.)

10.9 If X and Y are ideals of a ring, show that the *product* of X and Y,

$$X \cdot Y = \{x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n \quad | \quad x_i \in X \text{ and } y_i \in Y, \ n > 0\},$$

is an ideal.

10.10 In the ring of integers, find a positive integer n such that

$$(n) = (12) \cdot (16).$$

(See problem 10.9.)

10.11 Let X and Y be ideals of a ring. Prove that $X \cdot Y \subseteq X \cap Y$. (See problem 10.9.)

10.12 Let R be a ring and let p be a fixed prime. Define I_p to be the set of elements for which the order of the element is a power of p. Show that I_p is an ideal.

10.13 Find all of the subrings of the commutative ring of order 8 defined by tables 9.3 and 9.4 in chapter 9.

Hint: There are eight subgroups of the additive group Z_{15}^*. Find the eight subgroups, and determine which subgroups are in fact subrings.

10.14 Find all of the ideals of the commutative ring of order 8 defined by tables 9.3 and 9.4 in chapter 9. (See problem 10.13.)

10.15 Find all of the subrings of T_4 in table 9.5.

10.16 Find all of the ideals of T_4 in table 9.5.

10.17 Find all of the subrings of T_8 in table 9.6.

Hint: First find all 16 subgroups of the additive group, Z_{24}^*.

10.18 Find all of the ideals of T_8 in table 9.6. (See problem 10.17.)

10.19 Verify that $\{0, c\}$ is an ideal of the ring T_4 in table 9.5. Construct addition and multiplication tables for the quotient ring $T_4/\{0, c\}$.

10.20 Verify that $\{0, 2a\}$ is an ideal of the commutative ring R of order 8 which is defined by tables 9.3 and 9.4 in chapter 9. Construct addition and multiplication tables for the quotient ring $R/\{0, 2a\}$.

10.21 Verify that $\{0, b\}$ is an ideal of the commutative ring R of order 8 which is defined by tables 9.3 and 9.4 in chapter 9. Construct addition and multiplication tables for the quotient ring $R/\{0, b\}$.

10.22 Verify that $\{0, c\}$ is an ideal of the ring T_8 in table 9.6. Construct addition and multiplication tables for the quotient ring $T_8/\{0, c\}$.

10.23 Find a subring of the ring T_8 in table 9.6 that is isomorphic to the ring T_4 in table 9.5.

10.24 Determine all elements of T_8 in table 9.6 that have a multiplicative inverse.

10.25 Determine all elements of the ring defined by tables 9.3 and 9.4 in chapter 9 that have a multiplicative inverse.

10.26 An *irreducible* element p of a ring R is one for which the only way for $p = a \cdot b$ is for either a or b to have a multiplicative inverse. Determine the irreducible elements of the ring defined by tables 9.3 and 9.4 in chapter 9.

Hint: Cross out the rows and columns corresponding to the invertible elements. Which elements are no longer in the interior of the table?

10.27 Does T_4 or T_8 in tables 9.5 and 9.6 have any irreducible elements? (See problem 10.26.)

10.28 A *prime* element $p \neq 0$ of a ring R is a non-invertible element such that, whenever $a \cdot b$ is a multiple of p, either a or b is a multiple of p. (A multiple of p would be any element that can be expressed as either $x \cdot p$ or $p \cdot x$.) Find a prime element of the ring T_8 in table 9.6.

Hint: To determine if p is prime, first find all the multiples of p. Then cross out the rows and columns of the multiplication table corresponding to those elements. If there are no more multiples of p remaining, then p is prime.

10.29 Find a prime element of the ring defined by tables 9.3 and 9.4 in chapter 9 that is not irreducible. (See problems 10.26 and 10.28.)

10.30 Let R be a non-commutative ring. Define the operation $x * y = y \cdot x$. Show that the set R forms a ring using the operations $*$ and $+$ instead of $\cdot$ and $+$. This new ring is called the *transpose* of R, and is denoted R^t.

10.31 Show that the ring T_4 in table 9.5 is not isomorphic to its transpose. (See problem 10.30.)

10.32 Show that the ring T_8 in table 9.6 *is* isomorphic to its transpose. (See problem 10.30.)

Hint: First construct the multiplication table for T_8^t, then determine how to rearrange the elements of T_8 so that the patterns match.

10.33 Prove that a non-commutative ring of order 4 or less must be isomorphic to either T_4 from table 9.5 or T_4^t. (See problem 10.30.)

Hint: Use problem 9.22.

10.34 Is the ring $2\mathbb{Z}$ isomorphic to the ring $3\mathbb{Z}$? Why or why not?

10.35 Let $A = (6)$ be an ideal of the ring $\mathbb{Z}$. Construct addition and multiplication tables of the quotient ring $\mathbb{Z}/(6)$. What does this ring remind you of?

10.36 Let $A = (2)$ and $B = (6)$ be two ideals of the ring $\mathbb{Z}$. Construct addition and multiplication tables of the quotient ring A/B.

10.37 Let $A = (2)$ and $B = (8)$ be two ideals of the ring $\mathbb{Z}$. Show that the group A/B is isomorphic to Z_4, but the ring A/B is not isomorphic to the ring Z_4.

10.38 Find all ring homomorphisms from Z_6 to Z_6.

10.39 Show that if $\phi(x) = 2x$, then ϕ is *not* a ring homomorphism from $\mathbb{R}$ to $\mathbb{R}$.

10.40 Determine all ring homomorphisms from the rationals $\mathbb{Q}$ to $\mathbb{Q}$.
Hint: What are the possible kernels? If $\phi(1) = 1$, show that $\phi(n) = n$.

10.41 Let $\mathbb{C}$ denote the set of numbers of the form $a + bi$, where $i = \sqrt{-1}$ and a and b are real. ($\mathbb{C}$ is in fact a subring of the quaternions $\mathbb{H}$.) Let $\phi(a + bi) = a - bi$. Show that ϕ is a ring homomorphism from the ring $\mathbb{C}$ to itself.
Hint: Let $x = a + bi$, and $y = c + di$.

10.42 Let R be the extension of the ring $2Z_8 = \{0, 2, 4, 6\}$ by the integers. Find an ideal I of R such that $R/I \approx Z_8$.
Hint: Find a homomorphism from R onto Z_8, and use the first ring isomorphism theorem (10.2).

10.43 If R is a commutative ring and y is a fixed element of R, prove that the set

$$I = \{x \cdot y \quad | \quad x \in R\}$$

is an ideal of R.
Hint: Note that if there is no multiplicative identity, y may not be in I.

10.44 If R is a commutative ring and y is a fixed element of R, prove that the set

$$A = \{x \in R \quad | \quad x \cdot y = 0\}$$

is an ideal in R. (See problem 10.5.)

10.45 An element x of a ring R is called *nilpotent* if $x^n = 0$ for some positive number n. Show that the set of all nilpotent elements in a commutative ring R forms an ideal of R.
Hint: See problem 9.33.

10.46 Prove the second ring isomorphism theorem: If K and I are two ideals of a ring R, where $K \subseteq I$, then K is an ideal of I, I/K is an ideal of R/K, and

$$(R/K)/(I/K) \approx R/I.$$

10.47 Prove the third ring isomorphism theorem: If K and I are two ideals of a ring R, then

$$K/(K \cap I) \approx (K + I)/I.$$

(See problem 10.7 for the definition of $K + I$.)

This page intentionally left blank

Chapter 11

Integral Domains and Fields

11.1 Polynomial Rings

One major source of integral domains are the *polynomial rings.* We can construct a polynomial ring from any ring, but the polynomial rings with the familiar properties are formed either from fields or integral domains.

DEFINITION 11.1 Let K be a commutative ring. We define the set of polynomials in x over K, denoted $K[x]$, to be the set of all expressions of the form

$$k_0 + k_1x + k_2x^2 + k_3x^3 + \cdots$$

where the coefficients k_n are elements of K, and only a *finite* number of the coefficients are nonzero. If k_d is the last nonzero coefficient, then d is called the *degree* of the polynomial.

Notice that if $d = 0$, we essentially obtain the nonzero elements of K. These polynomials are referred to as *constant polynomials.* The degree for the zero polynomial

$$0 + 0x + 0x^2 + 0x^3 + \cdots$$

is not defined.

By convention, the terms with zero coefficients are omitted when writing polynomials. Thus, the second degree polynomial in $\mathbb{Z}[x]$

$$1 + 0x + 3x^2 + 0x^3 + \cdots$$

would be written $1 + 3x^2$. The one exception to this convention is the zero polynomial, which is written as 0.

We can define the sum and product of two polynomials in the familiar way. If

$$A = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots \qquad \text{and}$$
$$B = b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots$$

then

$$A + B = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + (a_3 + b_3)x^3 + \cdots$$

and

$$A \cdot B = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (a_i \cdot b_j) x^{(i+j)}.$$

Although this looks like a double infinite sum, only a finite number of the terms will be nonzero. If fact, this product could be written as

$$\begin{aligned} A \cdot B = a_0 \cdot b_0 & \\ &+(a_0 \cdot b_1 + a_1 \cdot b_0)x \\ &+(a_0 \cdot b_2 + a_1 \cdot b_1 + a_2 \cdot b_0)x^2 \\ &+(a_0 \cdot b_3 + a_1 \cdot b_2 + a_2 \cdot b_1 + a_3 \cdot b_0)x^3 + \cdots \end{aligned}$$

so each coefficient is determined by a finite sum.

LEMMA 11.1

Let A and B be two nonzero polynomials in x over K of degree m and n respectively, where K is a field or an integral domain. Then $A \cdot B$ is a polynomial of degree $m + n$, and $A + B$ is a polynomial of degree no greater than the larger of m or n.

PROOF Let A be a polynomial of degree m,

$$A = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots a_m x^m$$

and B be a polynomial of degree n,

$$B = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \cdots b_n x^n.$$

Here, a_m and b_n are nonzero elements of K. The product is determined by

$$A \cdot B = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i \cdot b_j \, x^{i+j}.$$

Note that a_i and b_j are zero for $i > m$ and $j > n$. If $i + j > m + n$, either $i > m$ or $j > n$, and in either case $a_i \cdot b_j = 0$. Thus, there are no nonzero terms in $A \cdot B$ with coefficients larger than $m + n$. However, if $i + j = m + n$, the only nonzero term would be the one coming from $i = m$ and $j = n$, giving

$$a_m b_n \, x^{m+n}.$$

Since there are no zero divisors in K, $a_m \cdot b_n$ is nonzero, so $A \cdot B$ is a polynomial of degree $m + n$.

Next we turn our attention to $A + B$. We may assume without loss of generality that m is no more than n. Then the sum of A and B can be expressed as

$$(a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots (a_m + b_m)x^m + b_{m+1}x^{m+1} + \cdots b_n x^n.$$

If $m < n$, this clearly is a polynomial with degree n. Even if $m = n$, this still gives a polynomial whose degree cannot be more than n. □

We still have to show that $K[x]$ will be a ring. But if K is an integral domain or field, we will be able to say more about $K[x]$.

PROPOSITION 11.1
Let K be an integral domain or a field. Then the set of polynomials in x over K forms an integral domain.

PROOF We have seen that $K[x]$ is closed under addition and multiplication. By the commutativity of K, addition and multiplication are obviously commutative. It is also clear that the zero polynomial acts as the additive identity in $K[x]$. Also, the additive inverse of

$$A = a_0 + a_1x + a_2x^2 + a^3x^3 + \cdots$$

is given by

$$-A = (-a_0) + (-a_1)x + (-a_2)x^2 + (-a_3)x^3 + \cdots,$$

since the sum of these two polynomials is

$$A + (-A) = 0 + 0x + 0x^2 + 0x^3 + \cdots = 0.$$

The polynomial with $b_0 = 1$, and $b_j = 0$ for all positive j,

$$I = 1 + 0x + 0x^2 + 0x^3 + \cdots,$$

acts as the multiplicative identity, since

$$I \cdot A = A \cdot I = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i \cdot b_j \, x^{i+j} = \sum_{i=0}^{\infty} a_i \cdot 1 \; x^i = A.$$

To check associativity of addition and multiplication, we need three polynomials

$$\begin{aligned} A &= a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots, \\ B &= b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots, \qquad \text{and} \\ C &= c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots. \end{aligned}$$

Then

$$\begin{aligned} (A + B) + C &= (a_0 + b_0) + c_0 + ((a_1 + b_1) + c_1)x + ((a_2 + b_2) + c_2)x^2 + \cdots \\ &= a_0 + (b_0 + c_0) + (a_1 + (b_1 + c_1))x + (a_2 + (b_2 + c_2))x^2 + \cdots \\ &= A + (B + C). \end{aligned}$$

Also,

$$\begin{aligned} A \cdot (B \cdot C) &= A \cdot \left(\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} b_j \cdot c_k \, x^{j+k} \right) \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_i \cdot (b_j \cdot c_k) x^{i+j+k} \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (a_i \cdot b_j) \cdot c_k \, x^{i+j+k} = (A \cdot B) \cdot C. \end{aligned}$$

The two distributive laws are also easy to verify using the summation notation.

$$\begin{aligned} A \cdot (B + C) &= A \cdot \left(\sum_{j=0}^{\infty} (b_j + c_j) x^j \right) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i \cdot (b_j + c_j) x^{i+j} \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (a_i \cdot b_j + a_i c_j) x^{i+j} \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i \cdot b_j \, x^{i+j} + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} a_i \cdot c_j \, x^{i+j} = A \cdot B + A \cdot C. \end{aligned}$$

We can use the fact that multiplication is commutative to show that $(A+B) \cdot C = A \cdot C + B \cdot C$. Thus, $K[x]$ is a commutative ring with identity.

Next, let us show that $K[x]$ has no zero divisors. Suppose that $A \cdot B = 0$, with both A and B being nonzero polynomials. Say that A has degree m and B has degree n. Then by lemma 11.1 $A \cdot B$ has degree $m + n$, which is impossible if either m or n were positive. But if A and B are constant polynomials, then $a_0 \cdot b_0 = 0$, which would indicate that either a_0 or b_0 is 0, since K has no zero divisors. Thus, either A or B would have to be 0, so we have that $K[x]$ has no zero divisors.

Finally, let us show that $K[x]$ is not a field, by showing that the polynomial $(1 + x)$ is not invertible. Suppose that there was a polynomial A such that $A \cdot (1 + x) = 1$. Then A is not 0. So suppose A has degree m. Then by lemma 11.1, we have $m + 1 = 0$, telling us $m = -1$, which is impossible. Thus, $(1+x)$ has no inverse in $K[x]$, and therefore $K[x]$ is an integral domain. □

Although this proposition holds for polynomials defined over a integral domain, there is no reason why we cannot have *Mathematica*® or GAP work with polynomials defined over any commutative ring. However, we will discover that the familiar properties of polynomials radically change!

Let us consider the commutative ring of order 8 from tables 9.3 and 9.4 in chapter 9.

InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, b]
Define[a.b, 0]; Define[b.a, 0]
R = Ring[{a, b}]

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,0],[0,b]]);
gap> List(R);
[ 0*a, b, a, a+b, 2*a, 2*a+b, 3*a, 3*a+b ]
```

We form a polynomial ring over R by defining a new symbol x. In GAP one uses the command `Indeterminate` to show that x is a variable over the ring R.

```
gap> x := Indeterminate(R,"x");
x
```

If there is a warning message, just ignore it. In *Mathematica* the symbol x is already available, but we must declare that x commutes with every element in R. This is done with the two definitions

Define[x.a, a.x]
Define[x.b, b.x]

which force the coefficients to appear in front of the power of x. A typical polynomial would be

Y = a.x + b

If we consider raising this polynomial to a power,

Y^4

we find that *Mathematica* writes the powers of x as $x.x.\ldots.x$. GAP does not have this difficulty:

```
gap> y:= a*x + b;
a*x+b
gap> y^4;
a*x^4+b
```

This polynomial ring has a rather bizarre property. Sometimes the square of a first degree polynomial is not a second degree polynomial! Consider

```
gap> (2*a*x + a+b)^2;
a+b
```

which yields the identity element in R. Furthermore, polynomials may be "factored" in more than one way. The two products

```
gap> (b+2*a*x)*(b+a*x);
2*a*x^2+b
gap> (b+2*a*x)*(2*a+b+a*x);
2*a*x^2+b
```

or, in *Mathematica*,

```
(b + 2 a.x) . (b + a.x)
(b + 2 a.x) . (2 a + b + a.x)
```

yield the same quadratic polynomial. Because of the bizarre properties of polynomials over general rings, we mainly will focus our attention to polynomial rings $K[x]$, where K is an integral domain or field.

As we work with polynomials in *Mathematica* we would like to use the standard multiplication notation instead of using the dot. There is a property of integral domains and fields that lets us enter these rings into *Mathematica* another way.

DEFINITION 11.2 Let R be a ring. We define the *characteristic of* R to be the smallest positive number n such that $n \cdot x = 0$ for all elements x of R. If no such positive number exists, we say the ring has *characteristic 0.*

PROPOSITION 11.2
Let R be a nontrivial ring without zero-divisors. If the characteristic is 0, then for n an integer and x a nonzero element of R, $n \cdot x = 0$ only if $n = 0$. If the characteristic is positive then it is a prime number p, and for nonzero x, $n \cdot x = 0$ if, and only if, n is a multiple of p.

PROOF Suppose that $n \cdot x = 0$ for some nonzero x in R. Then for any other nonzero element y of R,

$$0 = (n \cdot x) \cdot y = n \cdot (x \cdot y) = x \cdot (n \cdot y).$$

But x is nonzero, and the ring has no zero divisors, so we have $n \cdot y = 0$. This argument works in both ways, so

$$(*) \qquad n \cdot x = 0 \Longleftrightarrow n \cdot y = 0 \qquad \text{if } x \neq 0 \text{ and } y \neq 0.$$

If n was not zero, then $|n|$ would be a positive number such that $n \cdot x = 0$ for all x in the ring. Hence, if the ring has characteristic 0, then $n \cdot x = 0$ implies that either $x = 0$ or $n = 0$.

Now suppose that the ring has positive characteristic, and let x be any nonzero element of R. Let p be the smallest positive integer for which $p \cdot x = 0$. If p is not prime, then $p = a \cdot b$ with $0 < a < p$ and $0 < b < p$. But then

$$(a \cdot x) \cdot (b \cdot x) = (a \cdot b)\left(x^2\right) = (p \cdot x) \cdot x = 0 \cdot x = 0.$$

Since the ring has no zero divisors, either $a \cdot x = 0$ or $b \cdot x = 0$. But this contradicts the fact that p was the *smallest* number such that $p \cdot x = 0$. Thus, p is prime. By $(*)$ we have that $p \cdot y = 0$ for every element in R, and since this cannot be true for any smaller integer, we have that the characteristic of the ring is the prime number p.

It is easy to see that if n is a multiple of p, then $n = c \cdot p$ for some integer c. Thus, for any element x in R,

$$n \cdot x = (c \cdot p) \cdot x = c \cdot (p \cdot x) = c \cdot 0 = 0.$$

Suppose that $n \cdot x = 0$ for some n that is not a multiple of p. Then $\mathrm{GCD}(n, p)$ must be 1, and so by the greatest common divisor theorem (1.2), there are integers u and v such that $u \cdot n + v \cdot p = 1$. But then

$$x = 1 \cdot x = (u \cdot n + v \cdot p) \cdot x = u \cdot (n \cdot x) + v \cdot (p \cdot x) = u \cdot 0 + v \cdot 0 = 0.$$

So for nonzero x, $n \cdot x = 0$ if, and only if, n is a multiple of p. □

Characteristics are important because they provide a new way of defining integral domains and fields in *Mathematica*. We begin by telling *Mathematica* the characteristic p of the ring we want to define. For example, to define a ring with characteristic 3, we enter

InitDomain[3]

which does three things. First, it tells *Mathematica* that the ring to be defined is commutative, so the regular multiplication notation can be used instead of the dot. *Mathematica* defines the identity element to be 1. Finally, *Mathematica* assumes that the ring to be defined has no zero divisors, and takes into account proposition 11.2, defining three times *anything* to be 0. For example, the commands

2 + 2
2 i + 5 i

simplify to 1 and i. Let us try imitating the complex numbers, and tell *Mathematica* that $i^2 = -1$.

Define[i^2, −1]
K = Ring[{i}]
CheckRing[K]
AddTable[K]
MultTable[K]

This produces tables 11.1 and 11.2. We can define this ring in GAP as follows:

```
gap> InitRing("e","i");
gap> DefineRing("K",[3,3],[[e,i],[i,-e]]);
gap> CheckRing(K);
This is a ring.
```

TABLE 11.1: Addition of "complex numbers modulo 3"

$+$	0	1	2	i	$2i$	$1+i$	$2+i$	$1+2i$	$2+2i$
0	0	1	2	i	$2i$	$1+i$	$2+i$	$1+2i$	$2+2i$
1	1	2	0	$1+i$	$1+2i$	$2+i$	i	$2+2i$	$2i$
2	2	0	1	$2+i$	$2+2i$	i	$1+i$	$2i$	$1+2i$
i	i	$1+i$	$2+i$	$2i$	0	$1+2i$	$2+2i$	1	2
$2i$	$2i$	$1+2i$	$2+2i$	0	i	1	2	$1+i$	$2+i$
$1+i$	$1+i$	$2+i$	i	$1+2i$	1	$2+2i$	$2i$	2	0
$2+i$	$2+i$	i	$1+i$	$2+2i$	2	$2i$	$1+2i$	0	1
$1+2i$	$1+2i$	$2+2i$	$2i$	1	$1+i$	2	0	$2+i$	i
$2+2i$	$2+2i$	$2i$	$1+2i$	2	$2+i$	0	1	i	$1+i$

TABLE 11.2: Multiplication for "complex numbers modulo 3"

$\cdot$	0	1	2	i	$2i$	$1+i$	$2+i$	$1+2i$	$2+2i$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	i	$2i$	$1+i$	$2+i$	$1+2i$	$2+2i$
2	0	2	1	$2i$	i	$2+2i$	$1+2i$	$2+i$	$1+i$
i	0	i	$2i$	2	1	$2+i$	$2+2i$	$1+i$	$1+2i$
$2i$	0	$2i$	i	1	2	$1+2i$	$1+i$	$2+2i$	$2+i$
$1+i$	0	$1+i$	$2+2i$	$2+i$	$1+2i$	$2i$	1	2	i
$2+i$	0	$2+i$	$1+2i$	$2+2i$	$1+i$	1	i	$2i$	2
$1+2i$	0	$1+2i$	$2+i$	$1+i$	$2+2i$	2	$2i$	i	1
$2+2i$	0	$2+2i$	$1+i$	$1+2i$	$2+i$	i	2	1	$2i$

Unfortunately, this ring is just a little too large to display the addition and multiplication tables in GAP using the standard format (unless one resizes the screen). So we will number the elements to display the addition and multiplication tables.

```
gap> NumberElements := true;
true
gap> AddTable(K);

+       |1  2  3  4  5  6  7  8  9
--------+------------------------
0*e     |1  2  3  4  5  6  7  8  9
i       |2  3  1  5  6  4  8  9  7
2*i     |3  1  2  6  4  5  9  7  8
e       |4  5  6  7  8  9  1  2  3
e+i     |5  6  4  8  9  7  2  3  1
e+2*i   |6  4  5  9  7  8  3  1  2
2*e     |7  8  9  1  2  3  4  5  6
2*e+i   |8  9  7  2  3  1  5  6  4
2*e+2*i |9  7  8  3  1  2  6  4  5
```

```
gap> MultTable(K);

*       |1  2  3  4  5  6  7  8  9
--------+------------------------
0*e     |1  1  1  1  1  1  1  1  1
i       |1  7  4  2  8  5  3  9  6
2*i     |1  4  7  3  6  9  2  5  8
e       |1  2  3  4  5  6  7  8  9
e+i     |1  8  6  5  3  7  9  4  2
e+2*i   |1  5  9  6  7  2  8  3  4
2*e     |1  3  2  7  9  8  4  6  5
2*e+i   |1  9  5  8  4  3  6  2  7
2*e+2*i |1  6  8  9  2  4  5  7  3
```

Even with the abbreviated version of the multiplication table, we can see that this ring has nine elements and has no zero divisors. By corollary 9.1, K is a field. We could call K the field of "complex numbers modulo 3."

We can now form polynomials in K in *Mathematica* using the standard multiplication.

Y = (1 + i) x + 2;
Z = (2 + i) x^2 + 2 i x + 1 + 2i;
Y^2
$(2 + (1 + i)x)^2$
Y Z
$(2 + (1 + i)x)(1 + 2i + 2ix + (2 + i)x^2)$

Mathematica leaves the last two expressions in factored form. If we used the dot notation

Y.Y
$1 + x + ix + 2ix^2$
Y.Z
$2 + i + 2x + ix + 2x^2 + ix^2 + x^3$

instead, *Mathematica* expands the expressions. To do these same operations in GAP, we first define x to be an indeterminate in the ring K.

```
gap> x := Indeterminate(K,"x");
x
gap> y := (e+i)*x + 2*e;
(e+i)*x-e
gap> z := (2*e + i)*x^2 + 2*i*x + e + 2*i;
(2*e+i)*x^2+2*i*x+(e+2*i)
gap> y^2;
2*i*x^2+(e+i)*x+e
gap> y*z;
x^3+(2*e+i)*x^2+(2*e+i)*x+(2*e+i)
```

Mathematica and GAP can factor polynomials defined over any finite field. In the next chapter we will prove that such factorizations are unique. Even though the polynomial $x^2 + 1$ is irreducible over the integers, we can factor the polynomial over the field K:

Factor[x^2 + 1, K]

```
gap> Factor(x^2 + 1, K);
[ x+i, x+2*i ]
```

The polynomial rings defined over integral domains are the basic building blocks used for forming new integral domains and fields.

11.2 The Field of Quotients

In the last section, we found a way to form integral domains by imitating the familiar polynomials from high school algebra. In this section we will show how we can form a field from an integral domain, imitating grade school fractions.

We view a standard fraction as one integer divided by another. We want to extend this idea, and form fractions out of any integral domain. However, even with standard fractions there is a complication, since we consider

$$\frac{2}{4} = \frac{3}{6},$$

even though both the numerators and denominators are different. What we mean to say is that these two fractions are *equivalent*, where we define

$$\frac{x}{y} \equiv \frac{u}{v} \qquad \Leftrightarrow \qquad x \cdot v = y \cdot u.$$

This forms an equivalence relation on the set of fractions x/y. We have already seen equivalence relations while working with cosets of a group. What we call a rational number is really a set of fractions of the form x/y that are all equivalent.

DEFINITION 11.3 Let K be an integral domain, and let P denote the set of all ordered pairs (x, y) of elements of K, with y nonzero:

$$P = \{(x, y) \quad | \quad x, y \in K \quad \text{and} \quad y \neq 0\}.$$

We define a relation on P by

$$(x, y) \equiv (u, v) \qquad \text{if} \qquad x \cdot v = y \cdot u.$$

LEMMA 11.2
The above relation is an equivalence relation on P.

PROOF We need to show that the relation is reflexive, symmetric, and transitive. Let(x, y), (u, v), and (s, t) be arbitrary elements of P.

Reflexive:

$$(x, y) \equiv (x, y)$$

is equivalent to saying $x \cdot y = x \cdot y$ which is, of course, true. So this relation is reflexive.

Symmetric:

$$(x, y) = (u, v) \Longrightarrow x \cdot v = y \cdot u \Longrightarrow u \cdot y = v \cdot x \Longrightarrow (u, v) \equiv (x, y),$$

so this relation is also symmetric.

Transitive:

If $(x, y) \equiv (u, v)$ and $(u, v) \equiv (s, t)$, then

$$(x, y) \equiv (u, v) \Longrightarrow x \cdot v = y \cdot u \Longrightarrow x \cdot v \cdot t = y \cdot u \cdot t,$$

$$(u, v) \equiv (s, t) \Longrightarrow u \cdot t = v \cdot s \Longrightarrow u \cdot t \cdot y = v \cdot s \cdot y.$$

These two statements imply that $x \cdot v \cdot t = v \cdot s \cdot y$. Notice that in the last step we had to use the commutativity of multiplication. Using commutativity again, we have $x \cdot t \cdot v = y \cdot s \cdot v$, and since K has no zero divisors and v is nonzero, we can use lemma 9.4 to say that $x \cdot t = y \cdot s$. Then

$$x \cdot t = y \cdot s \Longrightarrow (x, y) \equiv (s, t),$$

so we have the transitive law holding. Therefore, this relation is an equivalence relation. □

DEFINITION 11.4 Let K be an integral domain, let P denote the set

$$P = \{(x, y) \quad | \quad x, y \in K \quad \text{and} \quad y \neq 0\},$$

and let the equivalence relation on P be

$$(x, y) \equiv (u, v) \qquad \text{if} \qquad x \cdot v = y \cdot u.$$

For each (x, y) in P, let $\left(\frac{x}{y}\right)$ denote the equivalence class of P that contains (x, y). Let Q denote the set of all equivalence classes $\left(\frac{a}{b}\right)$. The set Q is called the *set of quotients* for K.

This definition allows us to replace an equivalence of two expressions with an equality. We now have that

$$\left(\frac{x}{y}\right) = \left(\frac{u}{v}\right) \quad \text{if, and only if,} \quad x \cdot v = u \cdot y.$$

The next step is to define addition and multiplication on our set of quotients Q. Once again, we will use the rational numbers to guide us in the definition.

LEMMA 11.3
Let K be an integral domain, and let Q be the set of quotients for K. The addition and multiplication of two equivalence classes in Q, defined by

$$\left(\frac{x}{y}\right) + \left(\frac{u}{v}\right) = \left(\frac{x \cdot v + u \cdot y}{y \cdot v}\right)$$

and

$$\left(\frac{x}{y}\right) \cdot \left(\frac{u}{v}\right) = \left(\frac{x \cdot u}{y \cdot v}\right),$$

are both well defined operations on Q. That is, the sum and product do not depend on the choice of the representative elements (x, y) and (u, v) of the equivalence classes.

PROOF The first observation we need to make is that the formulas for the sum and product both form valid elements of Q, since $y \cdot v$ is nonzero as long as y and v are both nonzero.

Next let us work to show that addition does not depend on the choice of representative elements (x, y) and (u, v). That is, if $\left(\frac{x}{y}\right) = \left(\frac{a}{b}\right)$, and $\left(\frac{u}{v}\right) = \left(\frac{c}{d}\right)$, we need to show that

$$\left(\frac{x}{y}\right) + \left(\frac{u}{v}\right) = \left(\frac{a}{b}\right) + \left(\frac{c}{d}\right).$$

That is, we have to prove that

$$\left(\frac{x \cdot v + u \cdot y}{y \cdot v}\right) = \left(\frac{a \cdot d + c \cdot b}{b \cdot d}\right).$$

Since $\left(\frac{x}{y}\right) = \left(\frac{a}{b}\right)$ and $\left(\frac{u}{v}\right) = \left(\frac{c}{d}\right)$, we have $x \cdot b = a \cdot y$ and $u \cdot d = c \cdot v$. Multiplying the first equation by $v \cdot d$ and the second by $y \cdot b$, we get

$$x \cdot b \cdot v \cdot d = a \cdot y \cdot v \cdot d$$

and

$$u \cdot d \cdot y \cdot b = c \cdot v \cdot y \cdot b.$$

Adding this two equations together and factoring, we get

$$(x \cdot v + u \cdot y) \cdot b \cdot d = (a \cdot d + c \cdot b) \cdot y \cdot v.$$

This gives us

$$\left(\frac{x \cdot v + u \cdot y}{y \cdot v}\right) = \left(\frac{a \cdot d + c \cdot b}{b \cdot d}\right),$$

which is what we wanted.

We also need to show that multiplication is well defined, that is

$$\left(\frac{x}{y}\right) \cdot \left(\frac{u}{v}\right) = \left(\frac{a}{b}\right) \cdot \left(\frac{c}{d}\right).$$

But since $x \cdot b = a \cdot y$ and $u \cdot d = c \cdot v$, we can multiply these two equations together to get

$$x \cdot b \cdot u \cdot d = a \cdot y \cdot c \cdot v,$$

or

$$(x \cdot u) \cdot (b \cdot d) = (a \cdot c) \cdot (y \cdot v).$$

Therefore,

$$\left(\frac{x \cdot u}{y \cdot v}\right) = \left(\frac{a \cdot c}{b \cdot d}\right),$$

so multiplication also is well defined. □

THEOREM 11.1: The Field of Quotients Theorem
Let K be an integral domain, and let Q be the set of quotients for K. Then Q forms a field using the above definitions of addition and multiplication. The field Q is called the field of quotients for K.

PROOF We have already noted that addition and multiplication are closed in Q.

We next want to look at the properties of addition. From the definition,

$$\left(\frac{x}{y}\right) + \left(\frac{u}{v}\right) = \left(\frac{x \cdot v + u \cdot y}{y \cdot v}\right) = \left(\frac{u}{v}\right) + \left(\frac{x}{y}\right),$$

we see that addition is commutative. Let z be any nonzero element of K. Then $\left(\frac{0}{z}\right)$ acts as the additive identity:

$$\left(\frac{u}{v}\right) + \left(\frac{0}{z}\right) = \left(\frac{0}{z}\right) + \left(\frac{u}{v}\right) = \left(\frac{0 \cdot v + u \cdot z}{z \cdot v}\right) = \left(\frac{u \cdot z}{v \cdot z}\right) = \left(\frac{u}{v}\right).$$

Likewise, $\left(\frac{-u}{v}\right)$ is the additive inverse of $\left(\frac{u}{v}\right)$:

$$\left(\frac{u}{v}\right) + \left(\frac{-u}{v}\right) = \left(\frac{-u}{v}\right) + \left(\frac{u}{v}\right) = \left(\frac{-u \cdot v + u \cdot v}{v \cdot v}\right) = \left(\frac{0}{v \cdot v}\right) = \left(\frac{0}{z}\right).$$

The associativity of addition is straightforward:

$$\begin{aligned}\left(\left(\frac{x}{y}\right) + \left(\frac{u}{v}\right)\right) + \left(\frac{a}{b}\right) &= \left(\frac{x \cdot v + u \cdot y}{y \cdot v}\right) + \left(\frac{a}{b}\right) \\ &= \left(\frac{x \cdot v \cdot b + u \cdot y \cdot b + a \cdot y \cdot v}{y \cdot v \cdot b}\right),\end{aligned}$$

while

$$\begin{aligned}\left(\frac{x}{y}\right) + \left(\left(\frac{u}{v}\right) + \left(\frac{a}{b}\right)\right) &= \left(\frac{x}{y}\right) + \left(\frac{u \cdot b + a \cdot v}{v \cdot b}\right) \\ &= \left(\frac{x \cdot v \cdot b + u \cdot y \cdot b + a \cdot y \cdot v}{y \cdot v \cdot b}\right).\end{aligned}$$

So Q forms a group with respect to addition.

Next we look at the properties of multiplication. Multiplication is obviously commutative, since

$$\left(\frac{x}{y}\right)\cdot\left(\frac{u}{v}\right)=\left(\frac{x\cdot u}{y\cdot v}\right)=\left(\frac{u\cdot x}{v\cdot y}\right)=\left(\frac{u}{v}\right)\cdot\left(\frac{x}{y}\right).$$

We also have associativity for multiplication:

$$\begin{aligned}\left(\left(\frac{x}{y}\right)\cdot\left(\frac{u}{v}\right)\right)\cdot\left(\frac{a}{b}\right)&=\left(\frac{x\cdot u}{y\cdot v}\right)\cdot\left(\frac{a}{b}\right)\\&=\left(\frac{x\cdot u\cdot a}{y\cdot v\cdot b}\right)=\left(\frac{x}{y}\right)\cdot\left(\frac{u\cdot a}{v\cdot b}\right)=\left(\frac{x}{y}\right)\cdot\left(\left(\frac{u}{v}\right)\cdot\left(\frac{a}{b}\right)\right).\end{aligned}$$

The element $\left(\frac{z}{z}\right)$ acts as the multiplicative identity for any $z\neq 0$.

$$\left(\frac{z}{z}\right)\cdot\left(\frac{x}{y}\right)=\left(\frac{x}{y}\right)\cdot\left(\frac{z}{z}\right)=\left(\frac{x\cdot z}{y\cdot z}\right)=\left(\frac{x}{y}\right).$$

If $x=0$, then $\left(\frac{x}{y}\right)=\left(\frac{0}{z}\right)$. Otherwise, the multiplicative inverse of $\left(\frac{x}{y}\right)$ is $\left(\frac{y}{x}\right)$, since

$$\left(\frac{x}{y}\right)\cdot\left(\frac{y}{x}\right)=\left(\frac{x\cdot y}{y\cdot x}\right)=\left(\frac{z}{z}\right).$$

Thus, every nonzero element of Q has a multiplicative inverse. Finally, we have the two distribution laws. Because of the commutativity of multiplication, we only need to check one. Since

$$\left(\left(\frac{u}{v}\right)+\left(\frac{a}{b}\right)\right)\cdot\left(\frac{x}{y}\right)=\left(\frac{u\cdot b+a\cdot v}{v\cdot b}\right)\cdot\left(\frac{x}{y}\right)=\left(\frac{u\cdot b\cdot x+a\cdot v\cdot x}{v\cdot b\cdot y}\right),$$

while

$$\begin{aligned}\left(\frac{u}{v}\right)\cdot\left(\frac{x}{y}\right)+\left(\frac{a}{b}\right)\cdot\left(\frac{x}{y}\right)&=\left(\frac{u\cdot x}{v\cdot y}\right)+\left(\frac{a\cdot x}{b\cdot y}\right)\\&=\left(\frac{u\cdot x\cdot b\cdot y+a\cdot x\cdot v\cdot y}{v\cdot y\cdot b\cdot y}\right)\\&=\left(\frac{u\cdot x\cdot b+a\cdot x\cdot v}{v\cdot y\cdot b}\right),\end{aligned}$$

we have the distributive laws holding, and therefore Q is a field. □

In the construction of the field Q, we never used the identity element of K. Hence, if we started with a commutative ring without zero divisors instead of an integral domain, the construction would still produce a field. We can mention this as a corollary.

COROLLARY 11.1
Let K be any commutative ring without zero divisors. Then the set of quotients Q defined above forms a field.

Although the field of quotients was designed from the way we formed rational numbers from the set of integers, we can apply the field of quotients to any other integral domain. What happens if we form a field of quotients for the polynomial ring $K[x]$?

Let us first consider the most familiar polynomial ring $\mathbb{Z}[x]$—the polynomials with integer coefficients. An element in the field of quotients would be of the form $p(x)/q(x)$, where $p(x)$ and $q(x)$ are polynomials with integer coefficients. But we consider two such fractions $p(x)/q(x)$ and $r(x)/s(x)$ to be equivalent if $p(x) \cdot s(x) = r(x) \cdot q(x)$. For example, the two fractions

```
ClearDefs
A = (3 x^2 + 5 x - 2) / (2 x^2 + 7 x + 6)
B = (3 x^2 - 4 x + 1) / (2 x^2 + x - 3)
```

can be seen to be equivalent, since

```
Expand[(3 x^2 + 5 x - 2) * (2 x^2 + x - 3)]
Expand[(3 x^2 - 4 x + 1) * (2 x^2 + 7 x + 6)]
```

yield the same result. Other ways of showing that A and B are equivalent is by computing either of these two commands:

```
Simplify[A - B]
Simplify[A/B]
```

GAP has no problem is seeing that these rational functions are equivalent.

```
gap> x := Indeterminate(Integers,"x");
x
gap> A := (3*x^2 + 5*x - 2)/(2*x^2 + 7*x + 6);
(3*x-1)/(2*x+3)
gap> B := (3*x^2 - 4*x + 1)/(2*x^2 + x - 3);
(3*x-1)/(2*x+3)
```

We call the field of quotients for the polynomials $\mathbb{Z}[x]$ the *field of rational functions in* x, denoted $\mathbb{Z}(x)$.

It should be mentioned that a rational function, in this context, is not a function! The rational functions A and B are merely *elements* of $\mathbb{Z}(x)$, which may in turn be arguments for some homomorphism. To say that "A is undefined when $x = -2$" or "B is undefined at $x = 1$" is meaningless, since x is not a variable for which numbers can be plugged in. Rather, x is merely a symbol that is used as a place holder. This is why we can say that A and B are truly equal, even though the "graphs" would disagree at two points.

We can form rational functions from any integral domain K. This produces the field $K(x)$, the *rational functions in* x *over* K.

For example, let us use the field of order 9 that was defined by tables 11.1 and 11.2.

InitDomain[3]
Define[i^2, –1]
K = Ring[{i}]

Here is a typical rational function in x over K:

A = (i x^2 + x^2 + 2 x i + 2 x + 2)/(x^2 + i x + 1)

We can have *Mathematica* factor this over a finite field K with the command

Factor[A, K]

According to this factorization, the rational function A does not simplify. Or does it? Consider a simpler rational function.

B = (2 x – i) / (x – i x + i)
Simplify[A–B]

Mathematica shows us that these two expressions are the same rational function in $K(x)$. Again, GAP has no problem finding the simplification.

```
gap> InitRing("e","i");
gap> DefineRing("K",[3,3],[[e,i],[i,-e]]);
gap> x := Indeterminate(K,"x");
x
gap> A := (i*x^2 + x^2 + 2*x*i + 2*x + 2)/(x^2 + i*x +1);
((e+i)*x+(2*e+i))/(x+(e+2*i))
```

As you can see from this experiment, the definition of the quotient field does not depend on whether elements in the integral domain can be factored uniquely. However, unique factorization is an important property that we will study in depth in chapter 12.

11.3 Complex Numbers

We have already seen some examples of complex numbers in the form $a+bi$, where i represents the "square root of negative one." *Mathematica* uses a special blackboard i to display the imaginary number, but this can be entered into *Mathematica* as I. This allows us to perform standard arithmetic on complex numbers.

(2 + 3 I) + (4 – I)
$6 + 2\mathbf{i}$
(2 + 3 I) * (4 – I)
$11 + 10\mathbf{i}$
(2 + 3 I) / (4 – I)
$\frac{5}{17} + \frac{14\mathbf{i}}{17}$

GAP gives a more mysterious notation for the square root of -1:

```
gap> Sqrt(-1);
E(4)
```

The short explanation for this is that i is the fourth root of 1, that is, $i^4 = 1$. `E(4)` is GAP's notation for e_4, the principal fourth root of 1. Later in this section we will see how to find e_n for general n.

In spite of the unusual notation, we can still perform complex arithmetic in GAP, using `E(4)` for i.

```
gap> (2 + 3*E(4)) + (4 - E(4));
6+2*E(4)
gap> (2 + 3*E(4)) * (4 - E(4));
11+10*E(4)
gap> (2 + 3*E(4)) / (4 - E(4));
5/17+14/17*E(4)
```

In either presentation it is not at all clear where the "i" or e_4 came from. This gives the complex numbers a rather mysterious quality that is compounded by their common misnomer, "imaginary numbers."

Instead of considering quantities of the form $a+bi$, we will consider ordered pairs (a, b). We will declare the following properties for ordered pairs of real numbers:

1. $(a, b) = (c, d)$ if, and only if, $a = c$ and $b = d$.
2. $(a, b) + (c, d) = (a + c, b + d)$.
3. $(a, b) \cdot (c, d) = (a \cdot c - b \cdot d, a \cdot d + b \cdot c)$.

We define $\mathbb{C}$ to be the set of all ordered pairs of real numbers.

PROPOSITION 11.3

The set $\mathbb{C}$ forms a field, called the field of complex numbers. *This field contains a subfield isomorphic to the real numbers.*

PROOF Because the real numbers are closed with respect to both addition and multiplication, it is clear that both $(a+c, b+d)$ and $(a \cdot c - b \cdot d, a \cdot d + b \cdot c)$ would be defined for all real numbers a, b, c, and d. Thus, $\mathbb{C}$ is closed with respect to both addition and multiplication. Furthermore, since

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

and

$$(c, d) \cdot (a, b) = (c \cdot a - d \cdot b, c \cdot b + d \cdot a) = (a \cdot c - b \cdot d, a \cdot d + b \cdot c) = (a, b) \cdot (c, d),$$

we see that both addition and multiplication are commutative. The element $(0, 0)$ acts as the zero element, since

$$(0, 0) + (a, b) = (a, b).$$

The addition inverse of (a, b) is $(-a, -b)$, since

$$(a, b) + (-a, -b) = (0, 0).$$

Note that the order on the last two sums is irrelevant, since addition has already been shown to be commutative.

To show that addition is associative, we note that

$$(a, b) + \big((c, d) + (e, f) \big) = (a, b) + (c + e, d + f) = (a + c + e, b + d + f),$$

while

$$\big((a, b) + (c, d) \big) + (e, f) = (a + c, b + d) + (e, f) = (a + c + e, b + d + f).$$

To show that multiplication is associative is a little more complicated. We have

$$\begin{aligned}(a, b) \cdot \big((c, d) \cdot (e, f) \big) &= (a, b) \cdot (c \cdot e - d \cdot f, c \cdot f + d \cdot e) = \\ &(a \cdot c \cdot e - a \cdot d \cdot f - b \cdot c \cdot f - b \cdot d \cdot e, a \cdot c \cdot f + a \cdot d \cdot e + b \cdot c \cdot e - b \cdot d \cdot f),\end{aligned}$$

and

$$\begin{aligned}\big((a, b) \cdot (c, d) \big) \cdot (e, f) &= (a \cdot c - b \cdot d, a \cdot d + b \cdot c) \cdot (e, f) = \\ &(a \cdot c \cdot e - b \cdot d \cdot e - a \cdot d \cdot f - b \cdot c \cdot f, a \cdot c \cdot f - b \cdot d \cdot f + a \cdot d \cdot e + b \cdot c \cdot e).\end{aligned}$$

By comparing these two, we see that they are equal, so multiplication is associative.

We need to test the distributive laws next. The left distributive law we can get by expanding:

$$\begin{aligned}(a, b) \cdot \big((c, d) + (e, f) \big) &= (a, b) \cdot (c + e, d + f) \\ &= (a \cdot c + a \cdot e - b \cdot d - b \cdot f, a \cdot d + a \cdot f + b \cdot c + b \cdot e) \\ &= (a \cdot c - b \cdot d, a \cdot d + b \cdot c) + (a \cdot e - b \cdot f, a \cdot f + b \cdot e) \\ &= (a, b) \cdot (c, d) + (a, b) \cdot (e, f).\end{aligned}$$

Thus, the left distributive law is satisfied. However, the right distributive law follows from the left distributive law, and using the commutative multiplication:

$$\begin{aligned}\big((a, b) + (c, d) \big) \cdot (e, f) &= (e, f) \cdot \big((a, b) + (c, d) \big) \\ &= (e, f) \cdot (a, b) + (e, f) \cdot (c, d) \\ &= (a, b) \cdot (e, f) + (c, d) \cdot (e, f).\end{aligned}$$

We have now shown that the set $\mathbb{C}$ forms a commutative ring. To show that this ring has a multiplicative identity, we consider the element $(1, 0)$. Since the ring is commutative, we only need to check

$$(1, 0) \cdot (a, b) = (1 \cdot a - 0 \cdot b, 1 \cdot b + 0 \cdot a) = (a, b).$$

Finally, we need to show that every nonzero element has an inverse. If (a, b) is nonzero, then $a^2 + b^2$ will be a positive number. Hence

$$\left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right)$$

is an element of $\mathbb{C}$. The product

$$(a, b) \cdot \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right) = \left(\frac{a^2 + b^2}{a^2 + b^2}, \frac{-a \cdot b + a \cdot b}{a^2 + b^2} \right) = (1, 0)$$

verifies that

$$(a, b)^{-1} = \left(\frac{a}{a^2 + b^2}, \frac{-b}{a^2 + b^2} \right)$$

since multiplication is commutative. Therefore, the set $\mathbb{C}$ forms a field.

The second part of this proposition is to show that $\mathbb{C}$ contains a copy of the real numbers as a subfield. Consider the mapping f, which maps real numbers to $\mathbb{C}$, given by

$$f(x) = (x, 0).$$

To check that f is a homomorphism, we check that

$$f(x) + f(y) = (x, 0) + (y, 0) = (x + y, 0) = f(x + y)$$

and

$$f(x) \cdot f(y) = (x, 0) \cdot (y, 0) = (x \cdot y + 0, 0 + 0) = (x \cdot y, 0) = f(x \cdot y).$$

Thus, f is a homomorphism from the reals to $\mathbb{C}$. It is clear that f is one-to-one, since $(x, 0) = (y, 0)$ if, and only if, $x = y$. Thus, f is an embedding of the reals into $\mathbb{C}$, and thus the image of f:

$$\{(x, 0) \mid x \in \mathbb{R}\}$$

is isomorphic to the real numbers. □

LEMMA 11.4

There are exactly two solutions to the equation $x^2 = (-1, 0)$ *in the field* $\mathbb{C}$, *given by* $(0, \pm 1)$.

PROOF If (a, b) solves the equation $x^2 = (-1, 0)$, we have that

$$(a, b)^2 = (a^2 - b^2, 2a \cdot b) = (-1, 0).$$

Thus, a and b must satisfy the two equations

$$a^2 - b^2 = -1$$

and

$$2a \cdot b = 0.$$

The second equation implies that either a or b must be 0. But if $b = 0$, then the first equation becomes $a^2 = -1$, which has no real solutions. Thus, $a = 0$, and $-b^2 = -1$. There are two real solutions for b: ± 1. Thus, $(0, 1)$ and $(0, -1)$ both solve the equations for a and b, and so

$$(0, 1)^2 = (0, -1)^2 = (-1, 0).$$

□

We can now convert ordered pairs to the customary notation by defining $i = (0, 1)$, and identifying the identity element $(1, 0)$ with 1. Then any complex number (a, b) can be written

$$(a, b) = (a, 0) + (0, b) = a \cdot (1, 0) + b \cdot (0, 1) = a + bi.$$

We can rewrite the rules for addition and multiplication in $\mathbb{C}$ as follows:

$$(a + bi) + (c + di) = (a + c) + (b + d)i.$$

$$(a + bi) \cdot (c + di) = (a \cdot c - b \cdot d) + (b \cdot c + a \cdot d)i.$$

In working with groups, we found that the group automorphisms revealed many of the important properties of the group. This will also be true for rings. Let us extend the group automorphisms to apply to rings.

DEFINITION 11.5 A *ring automorphism* is a one-to-one and onto ring homomorphism that maps a ring to itself.

LEMMA 11.5
The set of all ring automorphisms of a given ring forms a group.

PROOF We first note that if $f(x)$ is an automorphism of a ring R, then $f^{-1}(x)$ is well defined, since $f(x)$ is both one-to-one and onto. We see that

$$f(f^{-1}(x) + f^{-1}(y)) = f(f^{-1}(x)) + f(f^{-1}(y)) = x + y,$$

so $f^{-1}(x + y) = f^{-1}(x) + f^{-1}(y)$. Also,

$$f(f^{-1}(x) \cdot f^{-1}(y)) = f(f^{-1}(x)) \cdot f(f^{-1}(y)) = x \cdot y,$$

so $f^{-1}(x \cdot y) = f^{-1}(x) \cdot f^{-1}(y)$. Thus, f^{-1} is a ring homomorphism. Since f was both one-to-one and onto, f^{-1} is both one-to-one and onto. Therefore, f^{-1} is a ring automorphism.

If f and ϕ are two ring automorphisms, then

$$f(\phi(x+y)) = f(\phi(x) + \phi(y)) = f(\phi(x)) + f(\phi(y))$$

and

$$f(\phi(x \cdot y)) = f(\phi(x) \cdot \phi(y)) = f(\phi(x)) \cdot f(\phi(y)).$$

The combination $f(\phi(x))$ is also one-to-one and onto, so this product, which we can denote $f \cdot \phi$, is a ring automorphism. Since the set of all ring automorphisms is closed with respect to multiplication and inverses, and the set of all ring automorphisms is a subgroup of the set of all *group* automorphisms with respect to addition, we see that this set is a group. □

The natural question that arises is determining all of the group of ring automorphisms of $\mathbb{C}$. This is in fact a difficult question to answer in general, but if we only consider the automorphisms that send each real number to itself, the question becomes easy to answer.

PROPOSITION 11.4

Besides the identity automorphism, there is another ring automorphism on $\mathbb{C}$*, given by*

$$\phi\big((a,b)\big) = (a,-b).$$

In fact, these are the only automorphisms for which $\phi(x) = x$ *for all real numbers* x*.*

PROOF We check that

$$\begin{aligned}\phi\big((a,b)\big) + \phi\big((c,d)\big) &= (a,-b) + (c,-d) = (a+c,-b-d)\\ &= \phi\big((a+c,b+d)\big) = \phi\big((a,b)+(c,d)\big).\end{aligned}$$

$$\begin{aligned}\phi\big((a,b)\big) \cdot \phi\big((c,d)\big) &= (a,-b)\cdot(c,-d) = (a\cdot c - b\cdot d, -a\cdot d - b\cdot c)\\ &= \phi\big((a\cdot c - b\cdot d, a\cdot d + b\cdot c)\big) = \phi\big((a,b)\cdot(c,d)\big).\end{aligned}$$

Thus, ϕ is a homomorphism. Since $(a,-b) = (0,0)$ if, and only if, a and b are both 0, the kernel of ϕ is just $\{(0,0)\}$, and so ϕ is one-to-one. Also, ϕ is onto, since $\phi\big((a,-b)\big) = (a,b)$. Therefore, ϕ is an automorphism.

To show that there are exactly two such automorphisms, suppose that $f(x)$ is an automorphism of $\mathbb{C}$ for which $f(x) = x$ for all real numbers x. Then $f\big((0,1)\big)^2 = f\big((0,1)^2\big) = f\big((-1,0)\big) = (-1,0)$, so by lemma 11.4 $f\big((0,1)\big) = (0,\pm 1)$. If $f\big((0,1)\big) = (0,1)$, then $f(x) = x$ for all $x \in \mathbb{C}$, and if $f\big((0,1)\big) = (0,-1)$, then $f(x) = \phi(x)$ for all x. □

The ring automorphism found in proposition 11.4 is called the *conjugate.* The conjugate of z is generally denoted by $\overline{z}$. That is, if $z = a + bi$, then $\overline{z} = \phi(z) = a - bi$. The conjugate automorphism is defined in *Mathematica* as

Conjugate[3 + 4 I]

or in GAP by

```
gap> ComplexConjugate(3 + 4*E(4));
3-4*E(4)
```

It is an easy computation to see that

$$z \cdot \overline{z} = (a + bi) \cdot (a - bi) = a^2 + b^2.$$

Thus, $z \cdot \overline{z}$ is always a non-negative real number.

DEFINITION 11.6 We say the *absolute value* of a complex number $z = a + bi$ is

$$|z| = \sqrt{z \cdot \overline{z}}.$$

The geometric interpretation of $|z|$ is the distance from (a, b) to the origin. In *Mathematica*, the function **Abs[z]** gives the absolute value for both real and complex numbers. There is no corresponding function in GAP, because GAP's square root function only works for rational numbers, and puts the answer in a nonstandard format.

PROPOSITION 11.5
For any two elements x and y in $\mathbb{C}$,

$$|x \cdot y| = |x| \cdot |y|.$$

PROOF We have

$$|x \cdot y| = \sqrt{x \cdot y \cdot \overline{x \cdot y}} = \sqrt{x \cdot y \cdot \overline{x} \cdot \overline{y}} = \sqrt{x \cdot \overline{x} \cdot y \cdot \overline{y}} = \sqrt{x \cdot \overline{x}} \cdot \sqrt{y \cdot \overline{y}} = |x| \cdot |y|.$$

Thus, $|x \cdot y| = |x| \cdot |y|$. □

From polar coordinates it is known that any point in the plane can be located by knowing its distance r from the origin, and its angle θ from the positive x-axis.

Since r is the absolute value of $(x+yi)$, perhaps the angle θ is also significant to the complex number. By using trigonometry in figure 11.1, we have that

$$x + yi = r(\cos\theta + i \sin\theta).$$

This form is called the *polar form* of the complex number $x + yi$. The angle θ is called the *argument* of $x + yi$. We can find the approximate argument of a complex number (in radians) with the *Mathematica* command

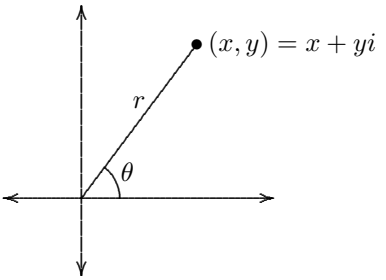


FIGURE 11.1: Polar coordinates for a complex number

N[Arg[3 + 4 I]]

Mathematica always finds an angle θ between $-\pi$ and π, but we can also consider the angles

$$\ldots,\ \theta - 6\pi,\ \theta - 4\pi,\ \theta - 2\pi,\ \theta,\ \theta + 2\pi,\ \theta + 4\pi,\ \theta + 6\pi,\ \ldots\ .$$

All of these angles have the same sine and cosine, and hence are interchangeable in the polar coordinate system. We call these angles *coterminal*. The set of angles coterminal to θ can be written

$$\{\theta + 2\pi n \mid n \in \mathbb{Z}\}.$$

For example, the polar form of $-\sqrt{3} - i$ is given by

$$2\left(\cos\left(\frac{-5\pi}{6}\right) + i\sin\left(\frac{-5\pi}{6}\right)\right),$$

as seen from the commands

Abs[– Sqrt[3] – I]
2
Arg[– Sqrt[3] – I]
$-\frac{5\pi}{6}$

However, we could have used any coterminal angle instead of the one *Mathematica* gave us. Thus,

$$2\left(\cos\left(\frac{7\pi}{6}\right) + i\sin\left(\frac{7\pi}{6}\right)\right), \qquad 2\left(\cos\left(\frac{19\pi}{6}\right) + i\sin\left(\frac{19\pi}{6}\right)\right), \quad \ldots$$

are also polar forms of $-\sqrt{3} - i$. The usefulness of the polar form of a complex number is hinted at by the next lemma, which makes use of the trigonometric identities

$$\begin{aligned} \cos(A + B) &= \cos(A)\cos(B) - \sin(A)\sin(B), \qquad \text{and} \\ \sin(A + B) &= \sin(A)\cos(B) + \cos(A)\sin(B). \end{aligned}$$

LEMMA 11.6
If $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ *and* $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$, *then*

$$z_1 \cdot z_2 = r_1 \cdot r_2\big(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\big).$$

So the argument of the product is the sum of the arguments.

PROOF We note that

$$z_1 \cdot z_2 = r_1(\cos\theta_1 + i\sin\theta_1) \cdot r_2(\cos\theta_2 + i\sin\theta_2) =$$
$$r_1 \cdot r_2((\cos\theta_1 \cdot \cos\theta_2 - \sin\theta_1 \cdot \sin\theta_2) + i \cdot (\cos\theta_1 \cdot \sin\theta_2 + \sin\theta_1 \cdot \cos\theta_2)).$$

Using the trigonometric identities, this simplifies to

$$z_1 \cdot z_2 = r_1 \cdot r_2\big(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)\big). \qquad \square$$

THEOREM 11.2: De Moivre's Theorem
If n *is an integer, and* $z = r(\cos\theta + i\sin\theta)$ *is a nonzero complex number in polar form, then*

$$z^n = r^n\big(\cos(n\theta) + i\sin(n\theta)\big).$$

PROOF Let us first prove the theorem for positive values of n. For $n = 1$, the statement is obvious. Let us assume that the statement is true for the previous case. That is,

$$z^{n-1} = r^{n-1}\big(\cos((n-1)\theta) + i\sin((n-1)\theta)\big).$$

We want to prove that the theorem holds for n as well. Using lemma 11.6, we have

$$\begin{aligned} z^n &= z^{n-1} \cdot z \\ &= r^{n-1}\big(\cos((n-1)\theta) + i\sin((n-1)\theta)\big) \cdot \big(r(\cos\theta + i\sin\theta)\big) \\ &= r^n(\cos((n-1)\theta + \theta) + i\sin((n-1)\theta + \theta)) \\ &= r^n(\cos(n\theta) + i\sin(n\theta)). \end{aligned}$$

Thus, the theorem is true for n, and hence by induction it is true whenever n is positive.

If z is nonzero, then letting $n = 0$ gives

$$r^0(\cos(0\,\theta) + i\sin(0\,\theta)) = 1(1 + i \cdot 0) = 1 = z^0.$$

So the theorem holds for $n = 0$. If z is nonzero, then $r > 0$, and so

$$\big(r^{-n}\big(\cos(-n\theta) + i\sin(-n\theta)\big)\big) \cdot \big(r^n\big(\cos(n\theta) + i\sin(n\theta)\big)\big) =$$
$$r^{-n+n}\big(\cos(-n\theta + n\theta) + i\sin(-n\theta + n\theta)\big) = r^0(\cos 0 + i\sin 0) = 1.$$

Now, if $n < 0$, then the theorem holds for $-n$, and so

$$z^{-n}\big(\, r^n(\cos(n\theta) + i\sin(n\theta))\,\big) = 1,$$

hence

$$r^n(\cos(n\theta) + i\sin(n\theta)) = z^n$$

even when $n < 0$. □

De Moivre's theorem (11.2) allows us to quickly raise a complex number to an integer power. For example, we can compute $(-\sqrt{3} - i)^5$ to be

$$2^5\left(\cos\left(\frac{-25\pi}{6}\right) + i\sin\left(\frac{-25\pi}{6}\right)\right) = 32\left(\frac{\sqrt{3}}{2} - \frac{i}{2}\right) = 16\sqrt{3} - 16i.$$

We can also use De Moivre's theorem (11.2) to find the n-th root of 1. We first define

$$e_n = \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right).$$

For example, $e_1 = 1$, $e_2 = -1$, $e_3 = (-1 + i\sqrt{3})/2$, and $e_4 = i$, which we have seen before. Then

$$(e_n)^n = \cos(2\pi) + i\sin(2\pi) = 1,$$

so e_n is indeed one n-th root of unity. In fact, all n-th roots of 1 are given by the numbers $e_n, e_n^2, e_n^3, \ldots$ up to $(e_n)^n = 1$.

Let us look at an example. The eighth root of unity, e_8, can be entered into *Mathematica* using the commands

```
InitDomain[0]
e8 = (1/2 + I/2) Sqrt[2]
```

The **InitDomain** command clears the previous fields that were defined, and allows us to use the dot for the product. This allows us to consider the group generated by e_8:

```
G = Group[{e8}]
```

This gives the eight roots of unity, and shows that these elements form a group. In fact, the n-th roots of unity will form a cyclic group isomorphic to Z_n.

By rearranging the elements of G, we can create a circle graph as in figure 11.2 with the elements in the proper positions in the complex plane.

```
G = { I, (1/2 + 1/2 I)Sqrt[2], 1 ,(1/2 – 1/2 I)Sqrt[2],– I,
(–1/2–1/2 I)Sqrt[2],–1,(–1/2+1/2 I)Sqrt[2]}
CircleGraph[G, Mult[e8]]
```

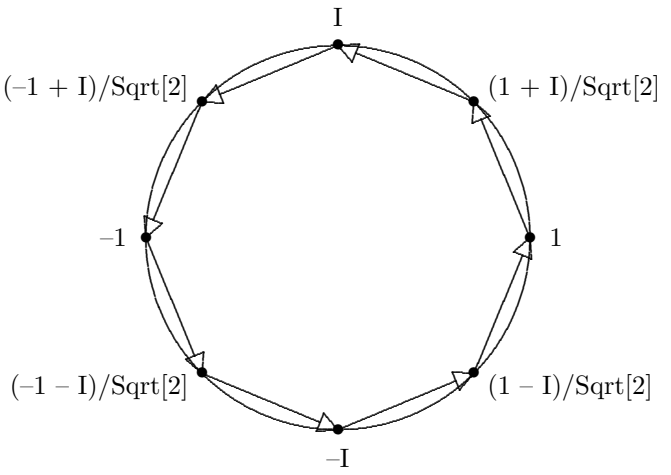


FIGURE 11.2: The eight roots of unity

We are mainly interested in those elements of this subgroup that are generators.

DEFINITION 11.7 A complex number z is called a *primitive n-th root of unity* if the powers of z produce all n solutions to the equation $x^n = 1$.

It is clear that e_n is a primitive n-th root of unity, but also $(e_n)^k$ is a primitive n-th root of unity if k and n are coprime.

We have already seen that GAP displays $\sqrt{-1}$ as e_4, which is a primitive root of unity, but GAP also calculates other square roots in terms of primitive roots of unity. Consider $\sqrt{2}$:

```
gap> Sqrt(2);
E(8)-E(8)^3
```

Since $e_8 = (1+i)/\sqrt{2}$ and $e_8^3 = (-1+i)/\sqrt{2}$, indeed $e_8 - e_8^3 = \sqrt{2}$. Here is a less obvious example.

```
gap> Sqrt(7);
E(28)^3-E(28)^11-E(28)^15+E(28)^19-E(28)^23+E(28)^27
```

Apparently $\sqrt{7}$ can be expressed in term of e_{28}. In fact, the square root of any rational number can be expressed in terms of some root of 1.

We have seen that we can use De Moivre's theorem 11.2 to raise a complex number to an integer power, or even a rational power. Is it possible to use this formula to raise a complex number to any real number, or even raise a number to a *complex* power?

In most fields, raising an element to the power of an *element* is absurd. Even in the real number system we will discover that we must utilize the exponential function e^x to compute quantities such as $2^{\sqrt{2}}$. We use that fact that $2 = e^{\ln 2}$, and so

$$2^{\sqrt{2}} = \left(e^{\ln 2}\right)^{\sqrt{2}} = e^{((\ln 2)\sqrt{2})}.$$

The key algebraic property of the exponential function is that

$$e^{x+y} = e^x \cdot e^y \qquad \text{for all} \quad x, y \in \mathbb{R}.$$

This indicates that the exponential function is a *group homomorphism* mapping the additive group of real numbers to the multiplicative group of real numbers. This homomorphism enables us to consider raising an element of the real numbers to the power of an *element*.

Can we extend the exponential function into a group homomorphism from the additive structure of $\mathbb{C}$ (denoted $\mathbb{C}^+$), to the multiplicative structure $\mathbb{C}^*$? If such a group homomorphism exists, then

$$e^{a+bi} = e^a \cdot e^{bi} = e^a \cdot (e^i)^b.$$

Mathematica indicates that the value of e^i is $(\cos 1 + i \sin 1)$. Problems 11.21 through 11.23 show three ways of proving this, all involving calculus. There is in fact no way to prove that $e^i = \cos 1 + i \sin 1$ without calculus. But given that this is true, we then have by De Moivre's theorem (11.2) that

$$e^{a+bi} = e^a \cdot (e^i)^b = e^a \cdot (\cos b + i \sin b)$$

whenever b is an integer. We will define this as the exponential function for all complex numbers. Notice that radian measure must be used in this formula.

PROPOSITION 11.6
For $z = a + bi$, the function

$$f(z) = e^a \cdot (\cos b + i \sin b)$$

defines a group homomorphism from $\mathbb{C}^+$ to $\mathbb{C}^*$, which is an extension of the standard exponential function. This function is called the complex exponential *function, and is also denoted e^z.*

PROOF If $z_1 = a_1 + b_1 i$, and $z_2 = a_2 + b_2 i$, we observe that

$$f(z_1 + z_2) = e^{a_1 + a_2}(\cos(b_1 + b_2) + i \sin(b_1 + b_2)).$$

By lemma 11.6, this equals

$$e^{a_1}(\cos(b_1) + i \sin(b_1)) \cdot e^{a_2}(\cos(b_2) + i \sin(b_2)) = f(z_1) \cdot f(z_2).$$

Thus, f is a group homomorphism from $\mathbb{C}^+$ to $\mathbb{C}^*$. □

This allows us another way of expressing e_n. Notice that

$$e^{2\pi i/n} = \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right) = e_n.$$

So we now have a more succinct way of defining the n-th root of 1.

The real exponential function is one-to-one, but is not onto since there is no number for which $e^x = -1$. However, the complex exponential function *is* onto, since for every nonzero complex number in polar form, $z = r(\cos\theta + i\sin\theta)$, there is a complex number whose exponential is z, namely $\ln(r) + i\theta$. The drawback of the complex exponential function is that it is *not* one-to-one! The kernel of this homomorphism is the set

$$N = f^{-1}(1) = \{2k\pi i \quad | \quad k \in \mathbb{Z}\}.$$

DEFINITION 11.8 For any nonzero complex number z, we define the *complex logarithm* of z, denoted $\log(z)$, to be the set of elements x such that $e^x = z$.

Notice that we use the function $\ln(x)$ to denote the *real* logarithm, while we use $\log(z)$ to denote the complex logarithm. We have already observed that when z is written in polar form, $z = r(\cos\theta + i\sin\theta)$, that one value of x that satisfies the equation is $x = \ln(r) + \theta i$. We also know that $f^{-1}(z)$ will be a coset of the kernel of f. Thus, we have $\log(z) = \ln(r) + \theta i + N$.

For example, $\log(-1)$ is the set

$$\{\pi i + 2k\pi i \quad | \quad k \in \mathbb{Z}\} = \{\ldots, -5\pi i, -3\pi i, -\pi i, \pi i, 3\pi i, 5\pi i, \ldots\}.$$

The *Mathematica* **Log** function works for complex numbers, but only gives one element of the set. Thus, we must add the kernel N to this result to obtain the set given by $\log(z)$.

We can now define a complex number raised to a complex power, by saying

$$x^z = (e^{\log(x)})^z = e^{z\cdot\log(x)}.$$

Notice that this gives a *set* of numbers, not just a single number. Although there will at times be an infinite number of elements in the set x^z, this will not always be the case.

PROPOSITION 11.7

For each integer $n > 0$, and any nonzero complex number z, then there are exactly n values for $z^{(1/n)}$. Thus, there are exactly n solutions for x to the equation $x^n = z$.

PROOF Let z have the polar form

$$z = r(\cos\theta + i\sin\theta).$$

Then $\log(z)$ is the set

$$\{\ln(r) + \theta i + 2k\pi i \mid k \in \mathbb{Z}\}.$$

Thus, $\log(z)/n$ is given by the set

$$\left\{ \frac{\ln(r)}{n} + \frac{(\theta + 2k\pi)i}{n} \;\middle|\; k \in \mathbb{Z} \right\}.$$

Thus, the exponential function of the elements of this set is given by

$$\left\{ e^{(\ln(r)/n} \cdot \left(\cos\left(\frac{(\theta + 2k\pi)}{n} \right) + i\sin\left(\frac{(\theta + 2k\pi)}{n} \right) \right) \;\middle|\; k \in \mathbb{Z} \right\}$$

$$= \left\{ r^{(1/n)} \cdot \left(\cos\left(\frac{(\theta + 2k\pi)}{n} \right) + i\sin\left(\frac{(\theta + 2k\pi)}{n} \right) \right) \;\middle|\; k \in \mathbb{Z} \right\}.$$

Notice that for two different values of k that differ by n, the arguments of the cosine and sine will differ by 2π. Hence, we only have to consider the values of k from 0 to $(n-1)$. This gives us the set

$$\left\{ r^{(1/n)} \cdot \left(\cos\left(\frac{(\theta + 2k\pi)}{n} \right) + i\sin\left(\frac{(\theta + 2k\pi)}{n} \right) \right) \;\middle|\; k = 0, 1, 2, \ldots, n-1 \right\}.$$

However, these n solutions will have arguments that differ by less than 2π so these n solutions are distinct.

Finally, we must show that x is an element of $z^{(1/n)}$ if, and only if, x solves the equation $x^n = z$. But for any element in the above expression, we have that

$$\begin{aligned} x^n &= r^{n(1/n)} \cdot \left(\cos\left(\frac{n(\theta + 2k\pi)}{n} \right) + i\sin\left(\frac{n(\theta + 2k\pi)}{n} \right) \right) \\ &= r(\cos\theta + i\sin\theta) = z. \end{aligned}$$

Likewise, if $x^n = z$, we can raise both sides to the $(1/n)$-th power to get that the two sets $(x^n)^{(1/n)}$ and $z^{(1/n)}$ are equal. Since the element x is certainly in the first set, it must also be in the set $z^{(1/n)}$ that we have just computed. □

This last proposition is very useful for finding square roots and cube roots of complex numbers. This turns out to have some important applications in finding the roots of real polynomials! In fact, complex numbers and the functions we have defined in this section also have many applications in the real world. The complex exponential function was fundamental to the invention of the short wave radio. The complex logarithm can be used in solving real valued differential equations. So even though these numbers are labeled "imaginary," they are by no means just a figment of someone's imagination.

11.4 Ordered Commutative Rings

The integers, the rational numbers, and the real numbers all have one property that most rings do not have. Given two different elements in the ring, we can say that one of them is greater than the other. Most rings do not have such an ordering, but we will find that some rings can be ordered in more than one way! The orderings of a ring can give us new insight into the structure of the ring.

We begin by making a formal definition of an ordered ring R. If there is a way to tell whether one element is greater than another, we should be able to distinguish those elements that are greater than zero, called the *positive elements* P.

DEFINITION 11.9 A commutative ring R is *ordered* if there exists a set P such that the three properties hold:

1. P is closed under addition.
2. P is closed under multiplication.
3. For each x in R, one and only one of the following statements is true:

$$x \in P, \qquad x = 0, \qquad -x \in P.$$

The third property is sometimes called the *law of trichotomy.* With this law, we can define what it means for one element to be greater than another.

DEFINITION 11.10 We say that x is greater than y, denoted $x > y$, if $x - y \in P$. Likewise, we say that x is smaller than y, denoted $x < y$, if $y - x \in P$. By the law of trichotomy, either

$$x > y, \qquad x < y, \qquad \text{or} \qquad x = y.$$

LEMMA 11.7
If x, y, and z are elements in an ordered ring, then we have the following three properties:

1. *If $x > y$, then $x + z > y + z$.*
2. *If $x > y$ and $z > 0$, then $x \cdot z > y \cdot z$.*
3. *If $x > y$ and $y > z$, then $x > z$.*

PROOF To prove the first statement, note that since $x > y$, we have that

$$x - y \in P.$$

But then

$$(x + z) - (y + z) \in P$$

and so $x + z > y + z$.

For the second statement, we have that $x > y$ and $z > 0$, and so $(x-y) \in P$ and $z \in P$. Since P is closed under multiplication, we have that

$$(x - y) \cdot z = x \cdot z - y \cdot z \in P,$$

and so $x \cdot z > y \cdot z$.

Finally, if $x > y$ and $y > z$, then both $x - y \in P$ and $y - z \in P$. Since P is closed under addition, we have that

$$(x - y) + (y - z) = x - z \in P,$$

and so $x > z$. □

Given a ring that has an ordering, one of the great challenges is determining the set of positive elements P. There are at least some elements that must be in P.

PROPOSITION 11.8
For any nonzero element x in an ordered ring, x^2 is in P.

PROOF Since x is nonzero, by the law of trichotomy either $x > 0$, or $-x > 0$. If $x > 0$ then

$$x^2 = x \cdot x > 0.$$

On the other hand, if $-x > 0$, then

$$x^2 = (-x) \cdot (-x) > 0.$$

Thus, in either case x^2 is in P. □

An immediate consequence of this is that if the ring has an identity e, then $e > 0$, since $e = e^2$. An additional statement can be proved if the ring is an integral domain.

COROLLARY 11.2
If R is an ordered integral domain with multiplicative identity 1, and n is any positive integer, then $n \cdot 1$ is in P. In particular, the characteristic of R must be 0.

PROOF Since $1^2 = 1$ we have from proposition 11.8 that $1 > 0$. Proceeding by induction, let us assume that $(n-1) \cdot 1 > 0$, and show that $n \cdot 1 > 0$. But this is easy, since

$$n \cdot 1 = (n-1) \cdot 1 + 1 \cdot 1 = (n-1) \cdot 1 + 1 > 0.$$

Thus, we have that $n \cdot 1 > 0$ for every positive number n. This immediately implies that the characteristic is zero, for if R had a positive characteristic p, then $p \cdot 1 = 0$, and we would have $0 > 0$, a contradiction. □

The standard examples of ordered rings are the integers, the rationals, and the real numbers. It should be noted that the complex numbers do *not* form an ordered ring, since $i^2 = -1 < 0$, and by proposition 11.8, any square must be positive.

Here is an very different example of an ordered integral domain. Consider all numbers of the form $x + y\sqrt{2}$, where x and y are integers. This forms a ring, since the product of any two such numbers yields a number of the same form. We will call this ring $\mathbb{Z}[\sqrt{2}]$, the ring formed by adjoining $\sqrt{2}$ to $\mathbb{Z}$. By proposition 9.3, this ring has no zero divisors, so this is an integral domain.

The standard ordering of $\mathbb{Z}[\sqrt{2}]$ would be to let P consist of all numbers that are positive when viewed as a real number. But let us try to find a nonstandard ordering of $\mathbb{Z}[\sqrt{2}]$. By corollary 11.2, the positive integers must be in P, but there is no way of proving that $\sqrt{2}$ is in P. Thus, we can consider an ordering where $-\sqrt{2} \in P$. We can determine whether any other element was in P or not in P. For example, $1 + \sqrt{2}$ would be negative, since

$$(1 + \sqrt{2}) \cdot (1 - \sqrt{2}) = -1 < 0,$$

and $1 - \sqrt{2}$ is the sum of two numbers in P, so this term is in P.

To see what is really going on in this example, it is helpful to look at the ring automorphisms, which were introduced in the last section. The automorphism of particular interest is as follows:

$$f : \mathbb{Z}[\sqrt{2}] \to \mathbb{Z}[\sqrt{2}],$$

$$f(x + y\sqrt{2}) = x - y\sqrt{2}.$$

This automorphism can be defined in *Mathematica*. We define the ring $\mathbb{Z}[\sqrt{2}]$ with the command

InitDomain[0]

Since *Mathematica* already knows that **Sqrt[2]** · **Sqrt[2]** is 2, this is all we need to define the ring in *Mathematica*. We now can define the homomorphism:

```
Homomorph[F]
Define[ F[Sqrt[2]] , - Sqrt[2] ]
CheckHomo[F,{ 1, Sqrt[2] }]
```

Since we are working with an infinite ring, we included only a basis for the ring instead of the whole ring as the second argument of the **CheckHomo** command.

To define this homomorphism in GAP, we need to first define $\mathbb{Z}[\sqrt{2}]$, of which GAP will express the elements in terms of e_8.

```
gap> K := Field(Sqrt(2));
NF(8,[ 1, 7 ])
gap> f:=AlgebraHomomorphismByImagesNC(K,K,[Sqrt(2)],[-Sqrt(2)]);
[ E(8)-E(8)^3 ] -> [ -E(8)+E(8)^3 ]
gap> Image(f, 2 + 3*Sqrt(2));
2-3*E(8)+3*E(8)^3
```

The `NC` (no check) version is needed to define this homomorphism because GAP has a problem showing that $e_8 - e_8^3$ generates the ring. Also, GAP's definition of the "ordering" of the elements is different than one would expect.

```
Sqrt(2) > 2
true
```

(In fact, GAP's inequalities treat all rationals smaller than irrationals.)

If we let P denote the set of positive elements using the "standard" ordering, and let P' be the set of positive elements under the unusual ordering we saw above, then $P' = f(P)$. In fact, for any automorphism ϕ on an ordered ring, we can construct an alternative way to order the ring by using $\phi(P)$ instead of P for the set of positive elements.

While we are working with the integral domain $\mathbb{Z}[\sqrt{2}]$ we might mention what happens if we consider the field of quotients of this ring. In fact, the resulting quotient field would be the set

$$x + y\sqrt{2}, \qquad x, y \in \mathbb{Q}.$$

Mathematica can check that multiplicative inverses exist for this set, with the command

CheckField[{**1, Sqrt**[**2**]}]

The argument of the **CheckField** command is a basis for the additive group. *Mathematica* finds that the inverse of **C**[**1**] **+ C**[**2**] **Sqrt**[**2**] is

$$\frac{C[1] - C[2]\sqrt{2}}{C[1]^2 - 2C[2]^2.}$$

We will call this field $\mathbb{Q}[\sqrt{2}]$.

The command **CheckField** not only verifies that a field is possible, but also defines all of the field operations into *Mathematica*. Thus the expression

1/(1 + Sqrt[**2**])

now simplifies to $\sqrt{2}-1$.

As one might guess from the `Field` command, it was really $\mathbb{Q}(\sqrt{2})$ that we defined earlier in GAP. Hence we can do divisions in this field.

```
gap> 1/(1 + Sqrt(2));
-1+E(8)-E(8)^3
```

In fact, we can do basic arithmetic over any combination of the e_n without having to define the field separately. The smallest field containing all roots of 1, that is, $\mathbb{Q}(e_3, e_4, e_5, e_6, e_7, \ldots)$, is called the *field of cyclotomics*.

The automorphism f that we discovered earlier on $\mathbb{Z}[\sqrt{2}]$ extends to an automorphism on $\mathbb{Q}[\sqrt{2}]$. Thus, the unusual ordering that we gave to $\mathbb{Z}[\sqrt{2}]$ extends to the field of quotients.

PROPOSITION 11.9

Let R be an ordered integral domain, with P the set of positive elements. Then if Q is the field of quotients on R, then the ordering on R can be extended in a unique way to an ordering on Q. That is, there is a unique set P' that forms an ordering on Q, with

$$p \in P \Rightarrow \left(\frac{p}{1}\right) \in P'.$$

PROOF We will begin by showing that the ordering is uniquely determined. Since for any p in P, we have

$$\left(\frac{1}{p}\right) \cdot \left(\frac{p}{1}\right) = \left(\frac{p}{p}\right) = \left(\frac{1}{1}\right) = 1 \in P,$$

$\left(\frac{1}{p}\right)$ must be considered to be positive in the new ordering. But then $\left(\frac{n}{p}\right)$ must be positive whenever n and p are in P. Thus P' contains at least those elements of the form $\left(\frac{n}{p}\right)$, where n and p are in P. Note that every nonzero element in the field of quotients Q must be of one of the four forms

$$\left(\frac{n}{p}\right), \left(\frac{-n}{p}\right), \left(\frac{n}{-p}\right), \left(\frac{-n}{-p}\right),$$

where n and p are in P. But the first and the last expressions are equivalent, and the middle two are also equivalent. Thus, for every nonzero element of Q, either that element or its negative is of the form $\left(\frac{n}{p}\right)$, with n and p in P. Thus, P' cannot contain any more elements besides those of the form $\left(\frac{n}{p}\right)$, and hence P' is uniquely determined.

Now, suppose we consider the set of elements P' that can be expressed in the form $\left(\frac{n}{p}\right)$, where n and p are in P. Does this form an ordering on Q? We have already seen that the law of trichotomy has already been demonstrated.

All we need to show is that P' is closed under addition and multiplication. But this is clear by looking at the formulas

$$\left(\frac{x}{y}\right) + \left(\frac{u}{v}\right) = \left(\frac{x \cdot v + u \cdot y}{y \cdot v}\right)$$

and

$$\left(\frac{x}{y}\right) \cdot \left(\frac{u}{v}\right) = \left(\frac{x \cdot u}{y \cdot v}\right).$$

Thus, P' forms an ordering on Q, and is an extension of the ordering P. □

What if we consider numbers of the form

$$x + y\sqrt[3]{2} + z\sqrt[3]{4}, \qquad x, y, z \in \mathbb{Q}?$$

We can define this field in *Mathematica* with the command

InitDomain[0]
CheckField[{1, 2^(1/3), 2^(2/3)}]

We may now perform operations in this field, such as

1/(1 + 2^(1/3) − 3 2^(2/3))

Unfortunately, $\sqrt[3]{2}$ is not in the field of cyclotomics, so we have to define this field in GAP using a totally different way. We let a be an element for which $a^3 = 2$, that is, a will be a root to the polynomial $x^3 - 2$.

```
gap> x := Indeterminate(Rationals,"x");
x
gap> K := FieldExtension(Rationals,x^3-2);
<algebraic extension over the Rationals of degree 3>
gap> a := PrimitiveElement(K);
(a)
gap> 1/(1 + a - 3*a^2);
(-7/87-17/87*a-4/87*a^2)
```

As one can see, $\sqrt[3]{2}$ is entered as a, and $\sqrt[3]{4}$ is entered as a^2.

This field does not have a nontrivial automorphism, since the only element in the field for which $x^3 = 2$ is $\sqrt[3]{2}$. Thus, an automorphism f on this field sends $\sqrt[3]{2}$ to itself, and hence $f(x) = x$ for all x in this field. It is not surprising, then, that this field does not have an unusual ordering, as the field $\mathbb{Z}[\sqrt{2}]$ did.

Let us look at one more example of a field with several possible ways of defining an ordering on the field. Consider the set S of numbers of the form

$$x + y\cos\left(\frac{\pi}{9}\right) + z\cos\left(\frac{2\pi}{9}\right), \qquad x, y, z \in \mathbb{Q}.$$

Using trigonometric identities we can multiply two such numbers together to get a number in the same form. This can be verified by the command

Expand[(x1 + y1 Cos[Pi/9] + z1 Cos[2 Pi/9])
*** (x2 + y2 Cos[Pi/9] + z2 Cos[2 Pi/9])]**

$$x1\,x2 + \frac{y1\,y2}{2} + \frac{y2\,z1}{4} + \frac{y1\,z2}{4} + \frac{z1\,z2}{2} + x2\,y1\,\mathrm{Cos}\left[\frac{\pi}{9}\right] + x1\,y2\,\mathrm{Cos}\left[\frac{\pi}{9}\right] + \frac{1}{2}y2\,z1\,\mathrm{Cos}\left[\frac{\pi}{9}\right] + \frac{1}{2}y1\,z2\,\mathrm{Cos}\left[\frac{\pi}{9}\right] + \frac{1}{2}z1\,z2\,\mathrm{Cos}\left[\frac{\pi}{9}\right] + \frac{1}{2}y1\,y2\,\mathrm{Cos}\left[\frac{2\pi}{9}\right] + x2\,z1\,\mathrm{Cos}\left[\frac{2\pi}{9}\right] + x1\,z2\,\mathrm{Cos}\left[\frac{2\pi}{9}\right] - \frac{1}{2}z1\,z2\,\mathrm{Cos}\left[\frac{2\pi}{9}\right]$$

We can have *Mathematica* check that this is a field.

InitDomain[0]
CheckField[{1, Cos[Pi/9], Cos[2 Pi/9]}]

This command allows us to simplify rather complex divisions.

1/(4 + 3 Cos[Pi/9] – 5 Cos[2 Pi/9])

$$\frac{2}{163}\left(45 - 58\mathrm{Cos}\left[\frac{\pi}{9}\right] + 48\mathrm{Cos}\left[\frac{2\pi}{9}\right]\right)$$

Since $\cos(\pi/9)$ can be expressed as $(e_{18} + 1/e_{18})/2$, and $\cos(2\pi/9) = (e_9 + 1/e_9)/2$, this field is a subfield of the field of cyclotomics.

```
gap> a := (E(18) + 1/E(18))/2;
-1/2*E(9)^4-1/2*E(9)^5
gap> b := (E(9) + 1/E(9))/2;
-1/2*E(9)^2-1/2*E(9)^4-1/2*E(9)^5-1/2*E(9)^7;
gap> K := Field(a);
NF(9,[ 1, 8 ])
gap> 1/(4 + 3*a - 5*b);
-48/163*E(9)^2-90/163*E(9)^3+10/163*E(9)^4+10/163*E(9)^5
 -90/163*E(9)^6-48/163*E(9)^7
```

Since the elements of this field are all real there is a natural ordering of the elements of S. Are there other ways to order this field? We want to look for automorphisms on the field S. But consider the following homomorphism:

Homomorph[F]
Define[F[Cos[Pi/9]], – Cos[2 Pi/9]]
Define[F[Cos[2 Pi/9]], Cos[Pi/9] – Cos[2 Pi/9]]
CheckHomo[F,{1, Cos[Pi/9], Cos[2 Pi/9]}]

```
gap> f := AlgebraHomomorphismByImagesNC(K,K,[a],[-b]);
[ -1/2*E(9)^4-1/2*E(9)^5 ] ->
[ 1/2*E(9)^2+1/2*E(9)^4+1/2*E(9)^5+1/2*E(9)^7 ]
gap> Image(f,b);
1/2*E(9)^2+1/2E(9)^7
gap> a-b;
1/2*E(9)^2+1/2E(9)^7
```

Notice that by defining $f(a) = -b$, we automatically get that $f(b) = a - b$ in GAP. Furthermore, we could consider the homomorphism $f^2(x) = f(f(x))$:

F[F[Cos[Pi/9]]]
F[F[Cos[2 Pi/9]]]

```
gap> f^2
[ -1/2*E(9)^4-1/2*E(9)^5,
  -1/4*E(9)^2-1/2*E(9)^3-1/4*E(9)^4-1/4*E(9)^5-1/2*E(9)^6
     -1/4*E(9)^7,
  -1/8*E(9)^3-3/8*E(9)^4-3/8*E(9)^5-1/8*E(9)^6 ] ->
[ -1/2*E(9)^2-1/2*E(9)^7,
  -1/2*E(9)^3+1/4*E(9)^4+1/4*E(9)^5-1/2*E(9)^6,
  -3/8*E(9)^2-1/8*E(9)^3-1/8*E(9)^6-3/8*E(9)^7 ]
gap> Image(f^2,a);
-1/2*E(9)^2-1/2E(9)^7
gap> Image(f^2,b);
1/2*E(9)^4+1/2*E(9)^5
```

This shows, among other things, that $f(f(a)) = b - a$ and $f(f(b)) = -a$. Are there any other automorphisms on the field S? We can show that this is all of them. We will take advantage of the trig identity $\cos(3x) = 4\cos^3 x - 3\cos x$.

Thus,

$$\frac{1}{2} = \cos\left(\frac{3\pi}{9}\right) = 4\cos^3\left(\frac{\pi}{9}\right) - 3\cos\left(\frac{\pi}{9}\right).$$

Thus, $\cos(\pi/9)$ satisfies the polynomial equation $4x^3 - 3x = 1/2$. Because f is an automorphism, we have to have $f(\cos(\pi/9))$ satisfying the same polynomial equation. But there are only three roots to a cubic equation, and so there are only three possible values for $f(\cos(\pi/9))$. Each of these three solutions produces a unique automorphism on S. By lemma 11.5, we see that the group of automorphisms of this ring is isomorphic to Z_3. The three automorphisms give us three ways to define an ordering on the field S:

1. $a >_1 b$ if a is larger than b as real numbers.
2. $a >_2 b$ if $f(a) >_1 f(b)$.
3. $a >_3 b$ if $f(f(a)) >_1 f(f(b))$.

Thus, we have seen that some fields may have many ways of assigning an order to the elements, while others have only 1. The key is the number of ring automorphisms. These ring automorphisms will play a major role in the following chapters.

Problems for Chapter 11

Interactive Problems

11.1 In the field of "complex numbers modulo 3":

InitDomain[3]
Define[i^2, –1]
CheckField[{1, i}]
K = Ring[{1, i}]

```
gap> InitRing("e","i");
gap> DefineRing("K",[3,3],[[e,i],[i,-e]]);
```

Factor the polynomials $x^3 + 1$, $x^3 + 2$, $x^3 + i$, $x^3 + 2i$. What do you notice about the factorizations? Knowing how *real* polynomials factor, explain what is happening.

11.2 Consider a rational function A in the field of "complex numbers modulo 3":

InitDomain[3]
Define[i^2, –1]
CheckField[{1, i}]
F = Ring[{1, i}]
A = (x^2 + x + i x + 2 + 2 i)/(x^2 + i x^2 + x + 2 i x + 1)
Factor[A, F]

Although A does not seem to simplify, there is a quotient of first degree polynomials that is equivalent to A. Find such a simplification.

Hint: Multiply the denominator by a constant so that the coefficient for the highest power of x is 1. Note that GAP would immediately find this simplification.

11.3 Follow the example of $\mathbb{Z}[\sqrt[3]{2}]$ to define the integral domain $\mathbb{Z}[\sqrt{5}]$ in *Mathematica* or GAP. Then define F to be a nontrivial ring automorphism for this domain.

11.4 Using the commands

InitDomain[0]
CheckField[{1, Cos[Pi/5] }]

verify that all numbers of the form $x + y\cos(\pi/5)$, where x and y are in $\mathbb{Q}$, form a field. Find a nontrivial ring automorphism on this field.

Hint: Use *Mathematica* to compute $\cos(\pi/5)$. How is this field related to the integral domain in problem 11.3?

11.5 Explain why the ring "complex numbers modulo 5":

InitDomain[5]
Define[i^2, –1]

```
gap> InitRing("e","i");
gap> DefineRing("F",[5,5],[[e,i],[i,-e]]);
```

does not form a field. Can you determine a pattern as to which integers "complex numbers modulo n" form a field?

11.6 Use GAP to calculate `Sqrt(5)` in terms of `E(5)`. Use this information to express $\sqrt{5}$ in terms of $\cos(2\pi/5)$ and $\cos(4\pi/5)$.

11.7 Use GAP to calculate `(Sqrt(17)-1)/4` in terms of `E(17)`. Use this information to express $(\sqrt{17}-1)/4$ in terms of $\cos(n\pi/17)$.

Non-Interactive Problems

11.8 Find the characteristic of the ring defined by tables 9.3 and 9.4 in chapter 9.

11.9 Find the characteristic of the ring T_8 in table 9.6.

11.10 Prove that if $n > 1$, the characteristic of Z_n is n.

11.11 Let R be a ring with identity. If the identity element has a finite order in the additive group, show that this order is the characteristic of the ring.

11.12 A *Boolean ring* is a nontrivial ring in which all elements x satisfy $x^2 = x$. Prove that every Boolean ring has characteristic 2.

11.13 Prove that if a ring R has a finite number of elements, then the characteristic of R is a positive integer.

11.14 If Q is the field of quotients of an integral domain, show that $\left(\frac{-a}{b}\right)$ is the additive inverse of $\left(\frac{a}{b}\right)$ in Q.

11.15 If Q is the field of quotients of an integral domain, show that the left distributive property holds for Q:

$$\left(\frac{u}{v}\right)\cdot\left(\left(\frac{x}{y}\right)+\left(\frac{z}{w}\right)\right)=\left(\frac{u}{v}\right)\cdot\left(\frac{x}{y}\right)+\left(\frac{u}{v}\right)\cdot\left(\frac{z}{w}\right).$$

11.16 If Q is the field of quotients of an integral domain, show that the multiplication in Q is associative.

11.17 Investigate what happens if we compute the field of quotients of a ring that is already a field. Let $K = Z_3$, and let P be the set of ordered pairs

$$P = \{(x, y) \quad | \quad x, y \in Z_3 \quad \text{and} \quad y \neq 0\}.$$

Write a list of all ordered pairs in P, and determine which pairs are equivalent under the relation

$$(x, y) \equiv (u, v) \quad \text{if} \quad x \cdot v \equiv y \cdot u \pmod 3.$$

If Q is the set of equivalence classes, construct addition and multiplication tables for Q and show that Q is isomorphic to Z_3.

11.18 Prove that if K is a field, then the field of quotients of K is isomorphic to K.

11.19 List all polynomials in $Z_3[x]$ that have degree 2.

11.20 Of the second degree polynomials in $Z_3[x]$ listed in problem 11.19, which ones cannot be factored?

Hint: A quadratic polynomial in $Z_3[x]$ cannot be factored if neither 0, 1, nor 2 are roots.

11.21 Assume that the Taylor series for the exponential function

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots$$

is valid for complex numbers as well as for real numbers. Prove that $e^i = (\cos 1 + i \sin 1)$.

Hint: Recall the Taylor series for $\sin(x)$ and $\cos(x)$.

11.22 Suppose we can write $e^{ix} = u(x) + iv(x)$, where $u(x)$ and $v(x)$ are real functions of a real variable x. If we assume that

$$\frac{d}{dx}e^{ix} = u'(x) + iv'(x) = ie^{ix},$$

use differential equations to prove that $u(x) = \cos(x)$ and $v(x) = \sin(x)$.

Hint: Since $e^0 = 1$, we know that $u(0) = 1$ and $v(0) = 0$.

11.23 Assume that the limit from calculus

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$$

is valid for complex values of x as well as real values. Prove that $e^i = (\cos 1 + i \sin 1)$.

Hint: Convert $(1 + i/n)$ into polar form using an arctangent.

11.24 Find all possible values of $\log(-1)$.

11.25 Find all possible values of $\log(\sqrt{3} - i)$.

11.26 Find all possible values of $1^{1/6}$.

11.27 Find all complex solutions to the equation $z^4 + 1 = 0$.

11.28 Find all complex solutions to the equation $z^3 + 8 = 0$

11.29 Find all possible values of $(8i)^{1/3}$.

11.30 Find five values of the expression i^i.

11.31 Find five values of the expression $(-i)^{(i/2)}$.

11.32 Show that when x and y are both complex, the set of all values of the expression x^y forms a geometric sequence:

$$\{\ldots, a \cdot r^{-3}, a \cdot r^{-2}, a \cdot r^{-1}, a, a \cdot r, a \cdot r^2, a \cdot r^3, \ldots\}.$$

11.33 Find complex numbers x and y such that the set of values for x^y are the powers of 2:

$$\{\ldots, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16, \ldots\}.$$

(See problem 11.32. There will be more than one solution to this problem.)

11.34 Show that for a fixed n, the set of all n-th roots of 1 forms a group with respect to multiplication.

11.35 Prove that the group in exercise 11.34 is cyclic, with

$$e_n = \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right)$$

as a generator. Show that any generator of this group is a primitive n-th root of unity.

11.36 Prove or disprove: For all complex numbers x, y, and z,

$$(x^z) \cdot (y^z) = (x \cdot y)^z.$$

Note: x^z and y^z may both represent *sets* of complex numbers, so the left hand side of this equation is the set of all possible products formed.

11.37 Prove or disprove: For all complex numbers x, y, and z,

$$(z^x)^y = z^{(x \cdot y)}.$$

(See the note on problem 11.36.)

11.38 Prove or disprove: For all complex numbers x, y, and z,

$$(z^x) \cdot (z^y) = z^{(x+y)}.$$

(See the note on problem 11.36.)

11.39 Show that the equation $x^2 + e = 0$ has no solutions in an ordered ring.

11.40 Prove that if a is an element in a nontrivial ordered ring, then there exists an element b such that $b > a$.

11.41 Prove that if x and y are two elements in an ordered ring,

$$x^2 + y^2 \geq 2xy.$$

11.42 Prove that if x and y are two elements in an ordered ring,

$$x^2 + y^2 \geq -2xy.$$

11.43 In the integral domain $\mathbb{Z}[x]$, let $(\mathbb{Z}[x])^+$ denote the set of all polynomials whose leading coefficient is positive. Prove that $\mathbb{Z}[x]$ is an ordered integral domain by proving that $(\mathbb{Z}[x])^+$ is a set of positive elements for $\mathbb{Z}[x]$.

11.44 Show that in the integral domain $\mathbb{Z}[x]$, there is a ring automorphism that sends x to $-x$. Hence, there is a second way to order the integral domain $\mathbb{Z}[x]$. Describe the set of positive elements in this new ordering. (See problem 11.43.)

11.45 Show that the ring of real numbers $\mathbb{R}$ does not have a nontrivial ring automorphism.

Hint: First show that there is no nonstandard ordering on $\mathbb{R}$.

Chapter 12

Unique Factorization

12.1 Factorization of Polynomials

Last chapter, we defined the integral domain $F[x]$ of all polynomials with coefficients in a field F. In this section we will investigate how such polynomials factor.

We say that $f(x)$ factors if there are two non-constant polynomials $g(x)$ and $h(x)$ such that $f(x) = g(x) \cdot h(x)$. We also say that both $g(x)$ and $h(x)$ divide the polynomial $f(x)$. But $g(x)$ and $h(x)$ may also factor into non-constant polynomials. We want to show that we can factor $f(x)$ into polynomials that cannot be factored further. We also want to lay down the groundwork for showing that the polynomials produced by this factorization are in some sense uniquely determined.

One of the standard techniques from a standard algebra course is doing "long division" on polynomials. A sample problem would be $x^3 - 3x^2 + 4x - 5$ divided by $2x^2 - 5$, done as figure 12.1. This shows that $x^3 - 3x^2 + 4x - 5$ divided by $2x^2 - 5$ yields $x/2 - 3/2$, with a remainder of $13/2x - 25/2$. We can write this as

$$x^3 - 3x^2 + 4x - 5 = (2x^2 - 5) \cdot (x/2 - 3/2) + (13/2x - 25/2).$$

$$\begin{array}{rl|l} & & x/2 \quad -3/2 \\ \hline 2x^2 - 5 & \big) \; x^3 - 3x^2 + 4x \quad -5 & \\ & \quad x^3 \qquad\quad -5/2x & \\ \hline & \quad -3x^3 + 13/2x - 5 & \\ & \quad -3x^3 \qquad\quad +15/2 & \\ \hline & & 13/2x - 25/2 \end{array}$$

FIGURE 12.1: Sample long division problem

Fortunately, *Mathematica*® and GAP can do this tedious long division for you.

```
PolynomialQuotient[x^3 - 3 x^2 + 4 x - 5, 2 x^2 - 5, x]
-3/2 + x/2
PolynomialRemainder[x^3 - 3 x^2 + 4 x - 5, 2 x^2 - 5, x]
-25/2 + 13 x/2
```

```
gap> x := Indeterminate(Rationals,"x");
gap> LongDivision(x^3 - 3*x^2 + 4*x - 5, 2*x^2 - 5);
[ 1/2*x-3/2, 13/2*x-25/2 ]
```

GAP makes a list of two polynomials, the first being the quotient, and the second the remainder. This "long division" algorithm works for any field, not just the rational numbers $\mathbb{Q}$. We can prove this by induction on the degree of the dividend.

THEOREM 12.1: The Division Algorithm Theorem
Let F be a field, and let $F[x]$ be the set of polynomials in x over F. Let $f(x)$ and $g(x)$ be two elements of $F[x]$, with g nonzero. Then there exist unique polynomials $q(x)$ and $r(x)$ in $F[x]$ such that

$$f(x) = g(x) \cdot q(x) + r(x)$$

and either $r(x) = 0$ or the degree of $r(x)$ is less than the degree of $g(x)$.

PROOF We begin by showing that $q(x)$ and $r(x)$ exist, and then prove that they are unique. If $f(x) = 0$, or if the degree of $f(x)$ is less than the degree of $g(x)$, we can simply let $q(x) = 0$, and $r(x) = f(x)$. So we may suppose that the degree of $f(x)$ is at least as large as the degree of $g(x)$. Let n be the degree of $f(x)$ and let m be the degree of $g(x)$.

If $n = m = 0$, then $f(x)$ and $g(x)$ are both nonzero constants in the field F, so we may pick $q(x)$ to be the constant polynomial $f \cdot g^{-1}$, and pick $r(x) = 0$. Thus, we can find a suitable $q(x)$ and $r(x)$ when $n = 0$.

Now let us proceed by induction on n. That is, we will assume that we can find a suitable $q(x)$ and $r(x)$ whenever the degree of $f(x)$ is less than n. Let

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0,$$

and

$$g(x) = b_m x^m + b_{m-1} x^{m-1} + \cdots + b_0.$$

Since n is at least as large as m, we can consider the polynomial

$$p(x) = a_n b_m^{-1} x^{n-m}$$

of degree $n - m$. By lemma 11.1, $p(x) \cdot g(x)$ has degree n, and in fact, since

$$p(x) \cdot g(x) = a_n x^n + a_n b_m^{-1} b_{m-1} x^{n-1} + \cdots + a_n b_m^{-1} b_0 x^{n-m},$$

the coefficient of the x^n term would be a_n. Thus, $f(x) - p(x) \cdot g(x)$ is of degree less than n. So by the induction hypothesis, there exist polynomials $z(x)$ and $r(x)$ such that

$$f(x) - p(x) \cdot g(x) = z(x) \cdot g(x) + r(x)$$

with the degree of $r(x)$, less than the degree of $g(x)$. Thus,

$$f(x) = (p(x) + z(x)) \cdot g(x) + r(x).$$

By letting $q(x) = p(x) + z(x)$ we have proved that suitable $q(x)$ and $r(x)$ exist.

Next, let us prove that $q(x)$ and $r(x)$ are unique. Suppose that there is a second pair $\overline{q}(x)$ and $\overline{r}(x)$ such that $f(x) = \overline{q}(x) \cdot g(x) + \overline{r}(x)$. Then

$$\overline{q}(x) \cdot g(x) + \overline{r}(x) = q(x) \cdot g(x) + r(x),$$

or

$$(\overline{q}(x) - q(x)) \cdot g(x) = r(x) - \overline{r}(x).$$

The left hand side is either 0 (when $\overline{q}(x) = q(x)$), or has degree at least m, since $g(x)$ is of degree m. The right hand side is either 0, or has a degree less than m. This is a contradiction unless both sides of the equation are 0. Thus, $\overline{q}(x) = q(x)$ and $\overline{r}(x) = r(x)$, and the uniqueness has been proven. □

This theorem not only shows that the quotient $q(x)$ and remainder $r(x)$ are unique, but the proof basically follows the procedure that it used in figure 12.1. This means that the familiar long division algorithm used for real polynomials will in fact work for polynomials over any field. In many circumstances, we can do this algorithm on polynomials over any integral domain.

COROLLARY 12.1

Let R be an integral domain, and let $f(x)$ and $g(x)$ be two polynomials in $R[x]$. If there is a field F containing R such that $g(x)$ divides $f(x)$ as polynomials in $F[x]$, and if the leading coefficient of $g(x)$ is 1, then $g(x)$ divides $f(x)$ in $R[x]$.

PROOF The only time that we needed to use a division in the proof of the division algorithm theorem (12.1) is when we divided by the leading coefficient of $g(x)$. Thus, if the leading coefficient of $g(x)$ is 1, we can do all of the operations in $R[x]$ instead of $F[x]$. The result is that there are polynomials $q(x)$ and $r(x)$ such that

$$f(x) = g(x) \cdot q(x) + r(x)$$

in $R[x]$. But $g(x)$ divides $f(x)$ in the ring $F[x]$. So there is an $h(x)$ in $F[x]$ such that

$$f(x) = g(x) \cdot h(x).$$

But $q(x)$ and $r(x)$ can also be viewed as polynomials in $F[x]$, and the division algorithm shows that these are uniquely defined, even in $F[x]$. Thus, $q(x) = h(x)$ and $r(x) = 0$. Therefore, $g(x)$ divides $f(x)$ in $R[x]$. □

We are used to thinking of polynomials as functions, rather than as elements in a domain. If we want to "evaluate" a polynomial $f(x)$ at a particular value y, we run into a technical problem, since $f(x)$ is not a function. The division algorithm comes to our rescue on the occasion when we do need to evaluate polynomials at a particular value.

DEFINITION 12.1 Let K be a field or integral domain, and let $K[x]$ be the set of polynomials in x over K. For a fixed element y in K, define the mapping $\phi_y : K[x] \to K$ by $\phi_y(f(x)) =$ the remainder $r(x)$ when $f(x)$ is divided by the first degree polynomial $(x - y)$. Since either $r(x)$ is 0 or is of degree 0, $r(x)$ is in fact in K.

PROPOSITION 12.1
The mapping $\phi_y : K[x] \to K$ is a homomorphism, called the evaluation homomorphism at y.

PROOF Let $f_1(x)$ and $f_2(x)$ be two polynomials in $K[x]$. By the division algorithm theorem (12.1) there exists $q_1(x)$, $q_2(x)$, $\phi_y(f_1(x)) = r_1(x)$, and $\phi_y(f_2(x)) = r_2(x)$ such that

$$f_1(x) = (x - y) \cdot q_1(x) + r_1(x),$$

and

$$f_2(x) = (x - y) \cdot q_2(x) + r_2(x).$$

Then

$$f_1(x) + f_2(x) = (x - y)(q_1(x) + q_2(x)) + r_1(x) + r_2(x),$$

and

$$\begin{aligned} f_1(x) \cdot f_2(x) &= ((x - y) \cdot q_1(x) + r_1(x)) \cdot ((x - y) \cdot q_2(x) + r_2(x)) \\ &= (x - y) \cdot ((x - y) \cdot q_1(x)q_2(x) + q_1(x)r_2(x) + q_2(x)r_1(x)) + r_1(x) \cdot r_2(x). \end{aligned}$$

By the uniqueness of the division algorithm, we have that

$$\phi_y(f_1(x) + f_2(x)) = r_1(x) + r_2(x) = \phi_y(f_1(x)) + \phi_y(f_2(x)),$$

and

$$\phi_y(f_1(x) \cdot f_2(x)) = r_1(x) \cdot r_2(x) = \phi_y(f_1(x)) \cdot \phi_y(f_2(x)).$$

Thus, ϕ_y is a homomorphism. □

We will often denote $\phi_y(f(x))$ by the conventional notation, $f(y)$. However, whenever we want to emphasize the homomorphism property, we will use the notation $\phi_y(f(x))$ for the evaluation homomorphism. In GAP, one can use the `Value` function to find the value of a polynomial in one variable at a particular number. To evaluate the polynomial $x^3 + 5x^2 + 4x - 4$ at $x = 3$, enter

```
gap> x := Indeterminate(Rationals,"x");
gap> Value(x^3 + 5*x^2 + 4*x - 4, 3);
80
```

This homomorphism is a bit more complicated in *Mathematica*. We can use the command **ReplaceAll**. This actually replaces every appearance of one symbol with another expression.

ReplaceAll[x^3 + 5 x^2 + 4 x − 4, x −> 3]

Notice how a minus sign and a greater than sign make up the arrow in this command. *Mathematica* also provides an abbreviation for this command:

x^3 + 5 x^2 + 4 x - 4 /. x −> 3

Here, the **/.** is an abbreviation for **ReplaceAll**, but it appears *after* the polynomial.

The `Value` and **ReplaceAll** commands suggest a way to determine what it means for a polynomial to have a root.

DEFINITION 12.2 Let $f(x)$ be a polynomial over the field or integral domain F. If r is an element of F such that $\phi_r(f(x)) = 0$, then r is called a *zero*, or a *root*, of $f(x)$. Of course this is equivalent to saying that $(x - r)$ is a factor of $f(x)$.

Example 12.1
Consider the polynomial $x^2 + 1$ in $Z_5[x]$. We can visually evaluate this polynomial at $x = 2$ to see that

$$\phi_2(x^2 + 1) = 2^2 + 1 = 0$$

in the field Z_5. Thus, 2 is a root, or zero, or $x^2 + 1$. □

As one can imagine, the factorization of a polynomial over an arbitrary field can be more cumbersome than the customary factorization. For a finite field (such as Z_5), almost the only way to find roots is by trial and error. Fortunately, *Mathematica* can do this very quickly. However, the good news is that if we have found enough roots to a polynomial, we already have the factorization.

PROPOSITION 12.2

Let $f(x)$ be a polynomial over the field F that has positive degree n and leading coefficient a_n. If $r_1, r_2, r_3, \ldots r_n$ are n distinct zeros of $f(x)$, then

$$f(x) = a_n \cdot (x - r_1) \cdot (x - r_2) \cdot (x - r_3) \cdots (x - r_n).$$

PROOF Again, we will proceed by induction on the degree of $f(x)$, which we will call n. If $n = 1$, then $f(x) = a_1x + a_0$, and since r_1 is a root, $a_1r_1 + a_0 = 0$. Thus, $a_0 = -a_1r_1$, and hence

$$f(x) = a_1x - a_1r_1 = a_1(x - r_1).$$

So the proposition is true when $n = 1$.

Now we will apply the induction hypothesis on n. Since r_n is a root of $f(x)$, we have that

$$f(x) = (x - r_n)g(x)$$

for some $g(x)$, which by lemma 11.1 is of degree $n-1$. Furthermore, $g(x)$ and $f(x)$ have the same leading coefficient, a_n. For $i = 1, 2, \ldots, n-1$, we have

$$0 = \phi_{r_i}(f(x)) = (r_i - r_n) \cdot \phi_{r_i}(g(x)).$$

Since $(r_i - r_n)$ is not 0, we have that $g(x)$ has $n-1$ distinct roots, namely r_1, r_2, r_3, $\ldots$, r_{n-1}. Thus, by induction,

$$g(x) = a_n(x - r_1)(x - r_2)(x - r_3) \cdots (x - r_{n-1}).$$

Thus,

$$f(x) = a_n(x - r_1)(x - r_2)(x - r_3) \cdots (x - r_n). \qquad \square$$

COROLLARY 12.2

A polynomial of positive degree n over the field F has at most n distinct zeros in F.

PROOF Suppose that $f(x)$ has at least $n + 1$ roots, $r_1, r_2, \ldots, r_n, r_{n+1}$. From proposition 12.2,

$$f(x) = a_n(x - r_1)(x - r_2)(x - r_3) \cdots (x - r_n).$$

Since r_{n+1} is also a root, we have

$$0 = \phi_{r_{n+1}}(f(x)) = a_n(r_{n+1} - r_1)(r_{n+1} - r_2)(r_{n+1} - r_3) \cdots (r_{n+1} - r_n).$$

But all of the terms on the right hand side are nonzero, which is a contradiction. Thus, there can be at most n distinct zeros of $f(x)$. $\square$

We can use proposition 12.2 to do some factorizations in different fields. For example, both 2 and 3 can be seen to be roots of the polynomial $x^2 + 1$ in $Z_5[x]$. Thus

$$x^2 + 1 = (x - 2)(x - 3) \quad \text{in} \quad Z_5.$$

Here is an application of corollary 12.2 that has many applications even using the real number field.

COROLLARY 12.3

Let F be a field, let $x_0, x_1, x_2, x_3, \ldots x_n$ be $n + 1$ distinct elements of F, and let $y_0, y_1, y_2, y_3, \cdots y_n$ be $n + 1$ values in F (not necessarily distinct). Then there is a unique polynomial $f(x)$ with degree at most n such that

$$f(x_0) = y_0, \quad f(x_1) = y_1, \quad f(x_2) = y_2, \quad \ldots \quad f(x_n) = y_n.$$

PROOF To prove uniqueness, suppose that $f(x)$ and $g(x)$ are two such polynomials. Then $h(x) = f(x) - g(x)$ will have roots at $x_0, x_1, x_2, x_3, \ldots, x_n$. But $h(x)$ would have degree at most n, which contradicts corollary 12.2. Thus, the polynomial $f(x)$ is unique.

To show that this polynomial exists, we will first construct the n-th degree polynomial

$$f_0(x) = \frac{(x - x_1) \cdot (x - x_2) \cdot (x - x_3) \cdots (x - x_n)}{(x_0 - x_1) \cdot (x_0 - x_2) \cdot (x_0 - x_3) \cdots (x_0 - x_n)}$$

for which $f_0(x_0) = 1$ but $x_1, x_2, x_3, \ldots x_n$ are roots of $f_0(x)$. (Note that since all of the x_i are distinct, the denominator is not 0.)

We can likewise define $f_1(x), f_2(x), f_3(x), \ldots, f_n(x)$ such that

$$f_1(x_1) = f_2(x_2) = f_3(x_3) = \cdots f_n(x_n) = 1,$$

yet the remaining n x_i's are roots for each polynomial. Finally, we construct the polynomial

$$g(x) = y_0 f_0(x) + y_1 f_1(x) + y_2 f_2(x) + y_3 f_3(x) + \cdots + y_n f_n(x).$$

Clearly $g(x)$ will be a polynomial of degree at most n, and also $g(x_0) = y_0$, $g(x_1) = y_1$, $g(x_2) = y_2$, $g(x_3) = y_3, \ldots g(x_n) = y_n$. Thus, we have constructed the required polynomial. ∎

This corollary shows, for example, that knowing just three points of a quadratic function is sufficient to determine the quadratic function. *Mathematica* and GAP have built-in functions that find this polynomial. For example, both the commands

InterpolatingPolynomial[{{1, 2}, {2, 4}, {3, 8}}, x]

```
gap> InterpolatedPolynomial(Rationals,[1,2,3],[2,4,8]);
x^2-x+2
```

find the polynomial in x such that $f(1) = 2$, $f(2) = 4$, and $f(3) = 8$. The format is slightly different in the two systems—in *Mathematica*, one gives a list of points, whereas in GAP one first lists the x values, and then the corresponding y values. Also the names of the functions are slightly different. Although this has the obvious applications to graphing polynomials, we will find in the next section some surprising real world applications when we apply this corollary to different fields.

We are now ready to define the polynomials that in many ways act as the prime numbers of number theory.

DEFINITION 12.3 A polynomial $f(x)$ in $F[x]$ is said to be *irreducible* over F if $f(x)$ has positive degree, and $f(x)$ cannot be expressed as a product $f(x) = g(x) \cdot h(x)$ where both $g(x)$ and $h(x)$ have positive degree. If $f(x)$ has positive degree and is not irreducible, it is called *reducible*.

We saw above that $x^2 + 1$ was reducible over Z_5. However, *Mathematica* and GAP will claim that this polynomial is irreducible.

Factor[x^2 + 1]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Factor(x^2 + 1, Rationals);
[ x^2 + 1 ]
```

The reason of course is that *Mathematica* and GAP are viewing this polynomial as an element of $\mathbb{Q}[x]$, not $Z_5[x]$. Yet this polynomial *does* have a factorization if we were allowed to work with complex numbers:

Expand[(x + I)(x − I)]

```
gap> (x + E(4))*(x - E(4));
x^2+1
```

Thus, x^2+1 is reducible over $\mathbb{C}$, the field of complex numbers. Thus, whether a polynomial is reducible or irreducible over F greatly depends on the field F.

It should be noted that if $g(x)$ and $h(x)$ both have positive degree, then $g(x) \cdot h(x)$ has degree at least 2. Thus, all polynomials of degree 1 must be irreducible. Constant polynomials, however, are not considered to be irreducible.

Although it can be tricky to decide whether a polynomial is reducible or irreducible, there is a way to test polynomials of low degree.

PROPOSITION 12.3

If $f(x)$ is a polynomial of degree 2 or 3 over the field F, then $f(x)$ is reducible over F if, and only if, $f(x)$ has a zero in F.

PROOF Suppose that $f(x)$ has a zero in F, say r. Then

$$f(x) = (x - r)q(x)$$

where $q(x)$ has degree one less than $f(x)$. This shows that $f(x)$ is reducible.

Now suppose that $f(x)$ is reducible. Then $f(x) = g(x) \cdot h(x)$, where the degree of $g(x)$ plus the degree of $h(x)$ is 2 or 3. Thus, either $g(x)$ or $h(x)$ has degree 1. We may suppose $g(x)$ has degree 1, and so

$$f(x) = (a_1x + a_0)h(x).$$

Then $-a_0a_1^{-1}$ is a root of $f(x)$, and the proof is complete. □

We can use this proposition to determine whether polynomials of degree less than 4 are irreducible over a finite field. Simply plug in all elements of the field, and see if any of them produce 0 in that field. For example, consider

$$x^3 + 2x^2 - 3x + 4 \qquad \text{over} \qquad Z_5.$$

We have:

```
x^3 + 2 x^2 - 3 x + 4 /. x -> 0
x^3 + 2 x^2 - 3 x + 4 /. x -> 1
x^3 + 2 x^2 - 3 x + 4 /. x -> 2
x^3 + 2 x^2 - 3 x + 4 /. x -> 3
x^3 + 2 x^2 - 3 x + 4 /. x -> 4
```

```
gap> x := Indeterminate(Rationals,"x");
gap> Value(x^3 + 2*x^2 - 3*x + 4, 0);
4
gap> Value(x^3 + 2*x^2 - 3*x + 4, 1);
4
gap> Value(x^3 + 2*x^2 - 3*x + 4, 2);
14
gap> Value(x^3 + 2*x^2 - 3*x + 4, 3);
40
gap> Value(x^3 + 2*x^2 - 3*x + 4, 4);
88
```

One of these, namely when x was replaced by 3, produced a multiple of 5, which is equivalent to 0 in the field Z_5. Thus, this polynomial is reducible.

PROPOSITION 12.4

If F is a field, then all polynomials in $F[x]$ of positive degree are either irreducible, or can be expressed as a product of irreducible polynomials.

PROOF If $f(x)$ has degree 1, then we have seen that it is irreducible. Let us proceed by induction on the degree n of $f(x)$. If $f(x)$ is not irreducible, then we can express $f(x) = g(x) \cdot h(x)$, where $g(x)$ and $h(x)$ are polynomials of degree at least 1. But $g(x)$ and $h(x)$ must have degree less than n. Thus, by induction, $g(x)$ and $h(x)$ are either irreducible, or can be written as a product of irreducible polynomials. Thus, $f(x)$ can be written as a product of irreducible polynomials. □

One last tool we have to help us find irreducible polynomials is the Greatest Common Divisor (GCD) of two polynomials. The proof of the next theorem mimics the proof of the greatest common divisor theorem for integers (1.2).

THEOREM 12.2: The Greatest Common Divisor Theorem for Polynomials

Let F be a field, and let $F[x]$ be the polynomials in x over the field F. Given two nonzero polynomials $f(x)$ and $g(x)$ in $F[x]$, there exists a nonzero polynomial $h(x)$ such that

1. *$h(x)$ divides both $f(x)$ and $g(x)$.*
2. *There exist polynomials $s(x)$ and $t(x)$ such that*

$$f(x) \cdot s(x) + g(x) \cdot t(x) = h(x).$$

Furthermore, the polynomial $h(x)$ is unique except for multiplication by a constant.

PROOF Let us consider the set of all polynomials that can be produced by

$$f(x) \cdot s(x) + g(x) \cdot t(x)$$

where $s(x)$ and $t(x)$ are in $F[x]$. Call this set A. Both $f(x)$ and $g(x)$ are in A, so A contains nonzero polynomials. Consider a nonzero polynomial $h(x)$ in A of the lowest degree. By the division algorithm theorem (12.1), we can find polynomials $q(x)$ and $r(x)$ such that

$$f(x) = q(x) \cdot h(x) + r(x),$$

where $r(x)$ is either 0, or has lower degree than $h(x)$. But then

$$r(x) = f(x) - q(x) \cdot h(x) = (1 - q(x) \cdot s(x)) \cdot f(x) - q(x) \cdot g(x) \cdot t(x),$$

which is in A. But if $r(x)$ is not zero, the degree of $r(x)$ would be less than the degree of $h(x)$, and we picked $h(x)$ to be of the lowest degree. Thus, $r(x) = 0$, and $h(x)$ divides $f(x)$. By a similar argument, $h(x)$ divides $g(x)$.

To prove that $h(x)$ is unique, note that since $h(x)$ divides $f(x)$ and $g(x)$, then $h(x)$ divides all polynomials in A. So if there is another polynomial $d(x)$ in A that divides both $f(x)$ and $g(x)$, then $h(x)$ would divide $d(x)$. But $d(x)$ would also divide $h(x)$. Thus, $h(x)$ and $d(x)$ would have to have the same degree, and

$$d(x) = u \cdot h(x)$$

where u is a constant polynomial. Thus, $h(x)$ is unique up to multiplication by a constant. □

DEFINITION 12.4 Given two polynomials in $F[x]$, the *greatest common divisor* is the polynomial given in the above theorem whose leading coefficient is 1.

The *Mathematica* command **PolynomialGCD** or GAP's `Gcd` will find the greatest common divisor of two polynomials. For example, GCD(x^4-1, x^3-1) is found by the commands

PolynomialGCD[x^3 - 1, x^4 - 1]

or the GAP command

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Gcd(x^3 - 1, x^4 - 1);
x-1
```

Thus, there are two polynomials $s(x)$ and $t(x)$ such that

$$(x^3 - 1) \cdot s(x) + (x^4 - 1) \cdot r(x) = x - 1.$$

COROLLARY 12.4
Let F be a field, and let $f(x)$, $g(x)$, and $h(x)$ be polynomials in $F[x]$. If $f(x)$ is an irreducible divisor of $g(x) \cdot h(x)$, then either $g(x)$ or $h(x)$ is a multiple of $f(x)$.

PROOF Suppose that $f(x)$ divides neither $g(x)$ nor $h(x)$. Then the greatest common divisor of $f(x)$ and $g(x)$ must have degree less than the degree of $f(x)$. But the GCD must divide $f(x)$, and $f(x)$ is irreducible. Thus, the greatest common divisor of $f(x)$ and $g(x)$ must be 1. Likewise the GCD of $f(x)$ and $h(x)$ must be also be 1. By the greatest common divisor theorem (12.2), there exist polynomials $r(x)$, $s(x)$, $t(x)$, and $u(x)$ such that

$$f(x) \cdot r(x) + g(x) \cdot s(x) = 1,$$

and

$$f(x) \cdot t(x) + h(x) \cdot u(x) = 1.$$

By multiplying these two together, we have

$$\begin{aligned} 1 &= (f(x) \cdot r(x) + g(x) \cdot s(x)) \cdot (f(x) \cdot t(x) + h(x) \cdot u(x)) \\ &= f(x)^2 \cdot r(x) \cdot t(x) + f(x) \cdot r(x) \cdot h(x) \cdot u(x) \\ &\qquad + f(x) \cdot g(x) \cdot s(x) \cdot t(x) + g(x) \cdot h(x) \cdot s(x) \cdot u(x). \end{aligned}$$

Note that all of the terms on the right hand side are multiples of $f(x)$ (including the last term, since $g(x) \cdot h(x)$ is a multiple of $f(x)$). But the left hand side is 1, which cannot be a multiple of $f(x)$. Thus, we have a contradiction, and so either $g(x)$ or $h(x)$ is a multiple of $f(x)$. □

The irreducible polynomials will play the same role in the domain $F[x]$ as prime numbers play in the domain $\mathbb{Z}$. The key property of integer factorizations is that every positive number greater than one can be factored *uniquely* into a product of primes. We would like to prove something similar for polynomials in $F[x]$, but find we will have to modify our definition of unique factorization. In the next section, we will explain what it means for a general ring to have a unique factorization, and apply this to both polynomial rings and integers.

12.2 Unique Factorization Domains

In this section we wish to determine a general definition of unique factorization that would apply not only to $F[x]$, but for *any* ring. We will mainly be interested in integral domains for which factorizations are unique.

DEFINITION 12.5 Let R be a commutative ring. We say that an element x in R is a *unit* if x has a multiplicative inverse.

In proposition 9.7 we defined the set of invertible elements of R as R^*, and showed that they formed a group under multiplication. The units of R will play the same role as the constant polynomials do in the ring $F[x]$. In fact, we can model the definition of reducible and irreducible elements of a ring on the definition of irreducible polynomials in $F[x]$.

DEFINITION 12.6 Let R be a commutative ring. If a nonzero element x in R is not a unit, and can be expressed as a product $x = y \cdot z$, where neither y nor z are units, then we say that x is *reducible*. If a nonzero element is neither a unit nor reducible, we say it is *irreducible*.

Although this definition is mainly applied to integral domains, we can apply the definition to any ring with an identity. Consider the ring defined by

tables 9.3 and 9.4 in chapter 9.

InitRing
Define[4 a, 0]; Define[2 b, 0]
Define[a.a, a]; Define[b.b, b]
Define[a.b, 0]; Define[b.a, 0]
R = Ring[{a, b}]

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,0],[0,b]]);
gap> ResetTableOptions();
gap> MultTable(R);
```

*	0*a	b	a	a+b	2*a	2*a+b	3*a	3*a+b
0*a	0*a	0*a	0*a	0*a	0*a	0*a	0*a	0*a
b	0*a	b	0*a	b	0*a	b	0*a	b
a	0*a	0*a	a	a	2*a	2*a	3*a	3*a
a+b	0*a	b	a	a+b	2*a	2*a+b	3*a	3*a+b
2*a	0*a	0*a	2*a	2*a	0*a	0*a	2*a	2*a
2*a+b	0*a	b	2*a	2*a+b	0*a	b	2*a	2*a+b
3*a	0*a	0*a	3*a	3*a	2*a	2*a	a	a
3*a+b	0*a	b	3*a	3*a+b	2*a	2*a+b	a	a+b

The units of this ring are $a+b$ and $3a+b$. But there is an irreducible element in this ring. Can you find it?

Let us consider the more familiar ring, $\mathbb{Z}$. The only two elements with multiplicative inverses are ± 1. The irreducible elements are of course the prime numbers 2, 3, 5, 7, 11, 13, ... But by this definition, the *negative* of a prime number is also irreducible. But by introducing negative primes, we find that numbers can be written as a product of primes in more than one way:

$$12 = 2 \cdot 2 \cdot 3 = 2 \cdot (-2) \cdot (-3) = (-2) \cdot (-2) \cdot 3.$$

Because we now are including negative primes, we also have to redefine what is meant by unique factorization. The first step is to understand the relationship between these different factorizations.

DEFINITION 12.7 Let R be a commutative ring with identity. We say that the element x is an *associate* of an element y if there is a unit z such that $y = x \cdot z$.

Note that if x is an associate of y, then $x = y \cdot z^{-1}$, so that y is an associate of x. Even though we saw three different factorizations of 12, note that these are related via associates. We now can explain what unique factorization means for a general ring.

DEFINITION 12.8 A ring R has *unique factorization* if the following two conditions are satisfied:

1. If x is nonzero, and is not a unit of R, then x can be written as a product of irreducible elements of R.

2. If
$$x = y_1 \cdot y_2 \cdot y_3 \cdots y_m = z_1 \cdot z_2 \cdot z_3 \cdots z_n$$
are two expressions of x as a product of irreducible elements, then $m = n$ and it is possible to reorder $z_1, z_2, \ldots z_n$ so that each pair (y_i, z_i) is associates.

Furthermore, if R is an integral domain, then R is a *unique factorization domain*, abbreviated as *UFD*.

We would like to find a quick way to determine whether an integral domain is a UFD. The needed tool will be the definition of the *prime* elements. Although we have already defined a prime element in the integers $\mathbb{Z}$, for a general ring we wish to define a prime element as one that satisfies a different property.

DEFINITION 12.9 A nonzero element x of a commutative ring is *prime* if x is not a unit, and whenever $y \cdot z$ is a multiple of x, then either y or z must be a multiple of x.

Although primes and irreducible elements are the same in $\mathbb{Z}$, for many other rings they are totally different. Consider the above ring of order 8. The irreducible element is also a prime element, but there are prime elements in this ring that are not irreducible. Can you find them? Although this ring has prime elements that are not irreducible, we can show that this can only happen when the ring has zero divisors.

LEMMA 12.1
If K is an integral domain, and x is a prime element of K, then x is irreducible.

PROOF Since x is prime, it is neither 0 nor a unit. Suppose that $x = y \cdot z$, where neither y nor z are units. Since x is prime, we have that either y or z is a multiple of x. Suppose that y is a multiple of x. Then $y = x \cdot w$ for some number w. Then
$$x = y \cdot z = x \cdot w \cdot z.$$
Since K is an integral domain, we know that x is not a zero divisor, so we can use lemma 9.3 and say that
$$1 = w \cdot z.$$

But this indicates that z is a unit, which contradicts the original assumption that neither y nor z were units. Thus, x is irreducible. ∎

Even though a prime element is irreducible in an integral domain, it is *not* true that an irreducible element is prime! Consider for example the ring $\mathbb{Z}[\sqrt{-5}]$, whose elements are the numbers of the form $x + y\sqrt{-5}$, where x and y are integers. To determine the irreducible elements of this ring, let us define the following function on $\mathbb{Z}[\sqrt{-5}]$:

$$N(x + y\sqrt{-5}) = (x + y\sqrt{-5})(x - y\sqrt{-5}) = x^2 + 5y^2.$$

Notice that $N(z)$ is the product of the number z with its complex conjugate. We can observe that if a and b are in $\mathbb{Z}[\sqrt{-5}]$, $N(a \cdot b) = N(a) \cdot N(b)$. This function will help us to determine the irreducible elements of $\mathbb{Z}[\sqrt{-5}]$.

Let us begin by finding the units of $\mathbb{Z}[\sqrt{-5}]$. If $a = x + y\sqrt{-5}$ is invertible, then $N(a)$ must be invertible. Hence $x^2 + 5y^2 = 1$. The only integer solution to this equation is when $y = 0$ and $x = \pm 1$. Thus, ± 1 are the two units of this ring.

Next, let us find an irreducible element. Since $N(2) = 4$, the only way a product of non-units a and b could equal 2 is if $N(a) = N(b) = 2$. But the equation $x^2 + 5y^2 = 2$ clearly has no integer solutions. Thus, 2 is an irreducible element in this ring. By the same reasoning, 3 is also irreducible.

However, neither 2 nor 3 is a prime element of this ring! Consider the product

$$(1 + \sqrt{-5})(1 - \sqrt{-5}) = 1 + 5 = 6.$$

This product is a multiple of 2 and 3, but neither factor is a multiple of 2 or 3. Thus, 2 and 3 are not prime in this ring.

This example shows a ring that is *not* a unique factorization domain. We have seen two ways of factoring the number 6 that are not equivalent in terms of associates. But the fact that neither 2 nor 3 is prime is a clue as to why this ring is not a UFD.

PROPOSITION 12.5

An integral domain is a UFD if, and only if, all nonzero, non-units can be written as a product of primes.

PROOF We begin be showing that if K is a UFD, then all irreducible elements are prime. Suppose w is irreducible, and $x \cdot w = y \cdot z$ is a multiple of w. Then x, y, and z have factorizations into irreducible elements:

$$x = x_1 \cdot x_2 \cdots x_n,$$

$$y = y_1 \cdot y_2 \cdots y_m,$$

$$z = z_1 \cdot z_2 \cdots z_k.$$

Thus,

$$x_1 \cdot x_2 \cdots x_n \cdot w = y_1 \cdot y_2 \cdots y_m \cdot z_1 \cdot z_2 \cdots z_k.$$

Since a factorization is unique, and all terms in this product are irreducible, we have that w is an associate to one of the terms on the right hand side. Thus, either y or z is a multiple of w, and hence w is prime.

Since a nonzero element that is not a unit in a UFD can be expressed as a product of irreducible elements, we have shown that all such elements can be expressed as a product of primes.

Now let us suppose that all nonzero, non-unit elements in an integral domain can be expressed as a product of primes. The first part of the definition of a UFD is obviously fulfilled since the prime elements are irreducible. Suppose we have another factorization in terms of irreducible elements.

$$p_1 \cdot p_2 \cdot p_3 \cdots p_n = z_1 \cdot z_2 \cdot z_3 \cdots z_m.$$

Here, the p_i are prime elements, while the z_j are merely irreducible elements. We need to prove that $n = m$, and that, after a rearrangement of the z_j's, we have that p_i and z_i are associates. We will proceed by induction on n, the number of primes in the factorization. If $n = 1$, then $m = 1$; otherwise we would have a prime number (which is irreducible) expressed as a product of two or more irreducible elements. Also, $p_1 = z_1$, and so trivially the P's are associates of the z's.

Next, we will consider the general case. Since the right hand side of

$$p_1 \cdot p_2 \cdot p_3 \cdots p_n = z_1 \cdot z_2 \cdot z_3 \cdots z_m$$

is a multiple of p_n, one of the z's must be a multiple of p_n. Suppose that

$$z_k = p_n \cdot u.$$

Since z_k is irreducible, we find that u is a unit, hence z_k and p_n are associates. We now can write

$$p_1 \cdot p_2 \cdot p_3 \cdots p_{n-1} \cdot p_n = z_1 \cdot z_2 \cdots z_{k-1} \cdot p_n \cdot u \cdot z_{k+1} \cdots z_m.$$

Since the ring is an integral domain, we can use lemma 9.3 and cancel out the p_n.

$$p_1 \cdot p_2 \cdot p_3 \cdots p_{n-1} = z_1 \cdot z_2 \cdots z_{k-1} \cdot (u \cdot z_{k+1}) \cdots z_m.$$

The unit u may be multiplied by any of the irreducible elements z to produce another irreducible element. We now can apply the induction hypothesis, which says that there are $n - 1$ z's left, and that a rearrangement of the z's would make p_i and z_i associates. Therefore, $m = n$, and some rearrangement of the z's in

$$p_1 \cdot p_2 \cdot p_3 \cdots p_n = z_1 \cdot z_2 \cdot z_3 \cdots z_m$$

will allow p_i and z_i to be associates, proving that the ring is a UFD. □

This proposition will help us greatly in determining whether an integral domain is a UFD. We usually will proceed in two steps: proving that any element can be written as a product of irreducible elements, and then proving that any irreducible element is prime.

COROLLARY 12.5
If F is a field, then the ring $F[x]$ is a UFD.

PROOF From proposition 12.4, every polynomial of positive degree is either irreducible, or can be expressed as a product of irreducible polynomials. By corollary 12.4, all irreducible polynomials are prime. Thus, by proposition 12.5, $F[x]$ is a UFD. □

Although this corollary proves that polynomials over the rational numbers have a unique factorization, we still have not proven that $\mathbb{Z}[x]$, the polynomials over the integers, is a unique factorization domain. Corollary 12.5 will not help us, since $\mathbb{Z}$ is not a field. Yet is seems plausible that we could prove that $\mathbb{Z}[x]$ is a UFD, merely by using the fact that $\mathbb{Q}[x]$ is a UFD. In the process, let us prove that $R[x]$ is a UFD whenever R is a UFD. First, we will need to prove a few lemmas. This next lemma, commonly referred to as Gauss' lemma, uses the formula for the product of two polynomials.

LEMMA 12.2: Gauss' Lemma
If R is an integral domain, then a prime element of R is also a prime element of $R[x]$.

PROOF We need to show that if p is a prime of R that divides $h(x) = f(x) \cdot g(x)$, then p must divide either $f(x)$ or $g(x)$. Suppose that p does not divide all of the coefficients of $f(x)$ nor does p divide all of the coefficients of $g(x)$. Let

$$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots,$$
$$g(x) = b_0 + b_1x + b_2x^2 + b_3x^3 + \cdots,$$
$$h(x) = f(x) \cdot g(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots.$$

Let a_i be the first coefficient of $f(x)$ that is not divisible by p, and let b_j be the first coefficient of $g(x)$ that is not divisible by p.

Since $h(x)$ is divisible by p, we know that the coefficient c_{i+j} must be divisible by p. But

$$c_{i+j} = a_0b_{i+j} + a_1b_{i+j-1} + \cdots + a_{i-1}b_{j+1} + a_ib_j + a_{i+1}b_{j-1} + \cdots a_{i+j}b_0.$$

Note that all terms on the right hand side except a_ib_j are divisible by p (since $a_0, a_1, \ldots a_{i-1}$ and $b_0, b_1, \ldots b_{j-1}$ are all multiples of p). So a_ib_j is also

a multiple of p. But this contradicts the fact that p is a prime element of R, and neither a_i nor b_j is a multiple of p. Thus, p is prime in $R[x]$. □

With Gauss' lemma (12.2), we can see that whenever a product of several polynomials in $R[x]$ is divisible by a p, a prime number of R, then one of those polynomials must have been divisible by p. We can use induction to extend this argument to any element of R.

LEMMA 12.3
Let R be a unique factorization domain, and let

$$g_1(x),\ g_2(x),\ g_3(x),\ \ldots,\ g_n(x)$$

be polynomials in $R[x]$ that are not divisible by any prime element of R. Let $f(x)$ be a polynomial in $R[x]$, and let c and d be two elements in R such that

$$c \cdot f(x) = d \cdot g_1(x) \cdot g_2(x) \cdot g_3(x) \cdots g_n(x).$$

Then d is divisible by c in R.

PROOF If c is a unit in R, then obviously d is a multiple of c. We will now use induction on the number of prime factors of c in the ring R. If c contains a prime p, then by lemma 12.2, one of the terms on the right hand side must be a multiple of p. But none of the $g_i(x)$ are divisible by a prime, so we find that d is a multiple of p. Then we have

$$\frac{c}{p} \cdot f(x) = \frac{d}{p} \cdot g_1(x) \cdot g_2(x) \cdot g_3(x) \cdots g_n(x),$$

where c/p and d/p are both in R. Since c/p contains one less prime factor than c, we can use induction to say that d/p is a multiple of c/p. Then d would be divisible by c in R. □

The next step in proving that $R[x]$ is a UFD is to find the irreducible elements of $R[x]$. If there is a field F that contains R, we can use the irreducible elements of $F[x]$ to find the irreducible elements of $R[x]$.

LEMMA 12.4
Let R be a unique factorization domain, and let F be a field containing R. Then if $f(x)$ is a polynomial in $R[x]$ that is irreducible in $F[x]$, then $f(x)$ can be written

$$f(x) = c \cdot g(x),$$

where c is an element of R, and $g(x)$ is irreducible in $R[x]$.

PROOF We want to first show that we can express

$$f(x) = c \cdot g(x),$$

where the only elements of R that divide $g(x)$ are units. Let a_0 be the constant coefficient of $f(x)$. Notice that if an element of R divides $f(x)$, then that element must divide a_0. Since R is a UFD, there are only a finite number of primes in the factorization of a_0. Let us proceed by induction on the number of primes in this factorization.

If there are no prime elements of R that divide $f(x)$ we can let $c = 1$ and $g(x) = f(x)$. If there *is* a prime element of R that divides $f(x)$, we can write

$$f(x) = p \cdot h(x),$$

where p is a prime in R, and $h(x)$ is in $R[x]$. But then the constant coefficient of $h(x)$ will contain one less prime in its prime factorization, so by induction we have

$$h(x) = d \cdot g(x),$$

where the only elements of R that divide $g(x)$ are units. Then we let $c = b \cdot d$, and

$$f(x) = c \cdot g(x).$$

All that is left to show is that $g(x)$ is irreducible in $R[x]$. Suppose that

$$g(x) = r(x) \cdot s(x),$$

where $r(x)$ and $s(x)$ are in $R[x]$. We then have

$$f(x) = c \cdot r(x) \cdot s(x).$$

But there is a field F containing R such that $f(x)$ is irreducible in $F[x]$. Thus, either $r(x)$ or $s(x)$ are units in $F[x]$, which are constant polynomials. But we designed $g(x)$ so that the only constants in $R[x]$ that divide $g(x)$ are units of R. Thus, $g(x)$ is irreducible in $R[x]$. □

Although this lemma refers to some field F that contains R, there is a natural field to use—the field of quotients in R. We can use this field to show that, in fact, the irreducible elements of R that we found in lemma 12.4 are in fact prime elements of $R[x]$.

LEMMA 12.5

Let R be a unique factorization domain, and let F be the field of quotients for R. Then if $g(x)$ is irreducible over $R[x]$ and $F[x]$, then $g(x)$ is prime in $R[x]$.

PROOF Suppose that $r(x) \cdot s(x)$ is divisible by $g(x)$ in $R[x]$. We need to show that either $r(x)$ or $s(x)$ is divisible by $g(x)$ in $R[x]$. Yet $g(x)$ is

irreducible in $F[x]$, which is a UFD since F is a field. Thus, either $r(x)$ or $s(x)$ is divisible by $g(x)$ in $F[x]$. Suppose that $r(x)$ is divisible. Then we have

$$r(x) = g(x) \cdot k(x),$$

where $k(x)$ is in $F[x]$. The coefficients of $k(x)$ are in the quotient field of R, so we may write

$$k(x) = \frac{a_0}{b_0} + \frac{a_1}{b_1}x + \frac{a_2}{b_2}x^2 + \frac{a_3}{b_3}x^3 + \cdots \frac{a_n}{b_n}x^n.$$

Let c be the product of $b_0 \cdot b_1 \cdot b_2 \cdot b_2 \cdots b_n$. Then $j(x) = c \cdot k(x)$ is an polynomial in $R[x]$. Thus we have

$$c \cdot r(x) = g(x) \cdot (c \cdot k(x)) = g(x) \cdot j(x),$$

where $g(x)$ and $j(x)$ are in $R[x]$. We now can apply lemma 12.4 to $j(x)$ and write

$$j(x) = d \cdot q(x),$$

where $q(x)$ is not divisible by any prime in R. Then

$$c \cdot r(x) = d \cdot g(x) \cdot q(x),$$

so we can apply lemma 12.3, since neither $g(x)$ nor $q(x)$ is divisible by a prime of R. Hence, d is divisible by c, and

$$r(x) = \frac{d}{c} \cdot g(x) \cdot q(x).$$

Therefore, $r(x)$ is divisible by $g(x)$, and hence $g(x)$ is prime in $R[x]$. □

At this point all of the major battles have been fought. All that is left to do is put the pieces together to show that $R[x]$ is UFD.

THEOREM 12.3: The Unique Factorization Domain Theorem
$R[x]$ is a unique factorization domain if, and only if, R is a unique factorization domain.

PROOF First of all, if R is not a UFD, then there is some element c of R that is not expressible as a product of primes. But then c cannot be expressed as a product of primes in $R[x]$, since such a product must consist of constant polynomials, and this would contradict the fact that c cannot be expressed as a product of primes in R. Thus, $R[x]$ would not be a UFD.

Now suppose that R is a UFD. We need to show that any nonzero polynomial $f(x)$ in $R[x]$ is either a unit, or is expressible as a product of prime polynomials. If $f(x)$ has degree 0, and is not a unit of R, then since R is a

UFD, the constant $f(x)$ can be expressed as a product of primes in R. By lemma 12.2, any prime in R is also a prime in $R[x]$. Thus, if the degree of $f(x)$ is zero, $f(x)$ is either a unit, or can be expressed as a product of primes in $R[x]$.

Now suppose $f(x)$ has positive degree. Let F be the field of quotients over R. Then $F[x]$ is a unique factorization domain by corollary 12.5. Thus, we can write

$$f(x) = g_1(x) \cdot g_2(x) \cdot g_3(x) \cdot \cdots \cdot g_n(x),$$

where each $g_i(x)$ is irreducible in $F[x]$. For each $g_i(x)$, let c_i be the product of the denominators of all of the coefficients. Then $h_i(x) = c_i \cdot g_i(x)$ will be in $R[x]$, and we have

$$\begin{aligned} c_1 \cdot c_2 \cdot c_3 \cdot \cdots \cdot c_n \cdot f(x) &= c_1 g_1(x) \cdot c_2 g_2(x) \cdot c_3 g_3(x) \cdot \cdots \cdot c_n g_n(x) \\ &= h_1(x) \cdot h_2(x) \cdot h_3(x) \cdot \cdots \cdot h_n(x). \end{aligned}$$

Since c_i is a unit in $F[x]$, the $h_i(x)$ will still all be irreducible in $F[x]$. We can now apply lemma 12.4 on each of the $h_i(x)$ and find an element d_i in R such that

$$h_i(x) = d_i \cdot j_i(x),$$

where the $j_i(x)$ are irreducible in $R[x]$. By lemma 12.5, the $j_i(x)$ are prime in $R[x]$. We now can express

$$c_1 \cdot c_2 \cdot c_3 \cdot \cdots \cdot c_n \cdot f(x) = d_1 j_1(x) \cdot d_2 j_2(x) \cdot d_3 j_3(x) \cdot \cdots \cdot d_n j_n(x).$$

Let $C = c_1 \cdot c_2 \cdot c_3 \cdots c_n$ and $D = d_1 \cdot d_2 \cdot d_3 \cdots d_n$. We can then write

$$C \cdot f(x) = D \cdot j_1(x) \cdot j_2(x) \cdot j_3(x) \cdots j_n(x),$$

where C and D are in R, and the $j_i(x)$ are prime polynomials in $R[x]$. We can now apply lemma 12.3, which states that D must be a multiple of C in R. Thus

$$f(x) = \frac{D}{C} \cdot j_1(x) \cdot j_2(x) \cdot j_3(x) \cdots j_n(x),$$

where D/C is in R. Since R is a UFD, D/C can be expressed as a product of primes in R, which by lemma 12.2 are primes in $R[F]$. Thus, $f(x)$ can be expressed as a product of primes in $R[x]$ and so by proposition 12.5, $R[x]$ is a UFD. □

Not only does this theorem determine when we can consider polynomial factorization to be unique, but this theorem also applies to factoring polynomials in more than one variable.

Since $R[x]$ is an integral domain, we can consider another variable y, and consider the polynomial ring $R[x][y]$. A typical element of $R[x][y]$ would be

$$c_0(x) + c_1(x)y + c_2(x)y^2 + c_3(x)y^3 + \cdots c_n(x)y^n,$$

where each $c_i(x)$ is a polynomial in $R[x]$. If each $c_i(x)$ is written

$$c_i(x) = d_0 + d_1x + d_2x^2 + d_3x^3 + \cdots$$

we find that the polynomial in $R[x][y]$ could be written

$$d_{0\,0} + d_{1\,0}x + d_{0\,1}y + d_{2\,0}x^2 + d_{1\,1}x \cdot y + d_{0\,2}y^2 + \cdots.$$

If we make the convention that $x \cdot y = y \cdot x$, we see that $R[x][y] = R[y][x]$.

DEFINITION 12.10 We will denote the polynomial ring of two variables by $R[x, y] = R[x][y]$. The variables x and y are called *indeterminates*. Likewise, we denote the polynomial ring of n indeterminates by

$$R[x_1, x_2, x_3, \ldots, x_n].$$

COROLLARY 12.6
Let R be a unique factorization domain and let $x_1, x_2, x_3, \ldots x_n$ be indeterminates over R. Then $R[x_1, x_2, x_3, \ldots x_n]$ is a unique factorization domain.

PROOF We will use induction on n. If $n = 1$, the unique factorization domain theorem (12.3) shows that $R[x]$ is a UFD. Otherwise, we write

$$R[x_1, x_2, x_3, \ldots, x_n] = R[x_1, x_2, x_3, \ldots, x_{n-1}][x_n].$$

By the induction hypothesis, $R[x_1, x_2, x_3, \ldots, x_{n-1}]$ is a UFD. So by the unique factorization domain theorem (12.3), $R[x_1, x_2, x_3, \ldots, x_n]$ is a UFD. □

Polynomials in several variables are of considerable importance in geometry, since curves and surfaces are described by equations in several variables. Although *Mathematica*'s **Factor** command will be able to factor polynomials in many variables, its ability is limited to when R is either $\mathbb{Z}$ or $\mathbb{Q}$. For example, *Mathematica* can factor

```
ClearDefs
Factor[x^3 y^2 + x^2 y - x y^2 - 2 x + y]
```

over the integers, but cannot factor this over any other ring, even a finite field. Yet we will not have a need for factoring polynomials in two variables over any other field.

GAP's ability to factor polynomials in two variables is still in development. A preview of the multivariable factorization package is included in the file "multivar.g" in the `gap` directory. In GAP 4.4.12, we must first read in this extra package before the multivarible factorization will work. This package will probably be included in future versions of GAP.

```
gap> Read("c:/gap/multivar.g");
gap> x := Indeterminate(Rationals,"x");
x
gap> y := Indeterminate(Rationals,"y");
y
Factors(x^3*y^2 + x^2*y - x*y^2 - 2*x + y);
[ x*y-1, x^2*y+2*x-y ]
```

12.3 Principal Ideal Domains

Although we have found that polynomial rings created from unique factorization domains produce more unique factorization domains, there still is the question of how to tell whether a given ring is a unique factorization domain. The answer lies in the ideals of the ring. In fact, the ideals were discovered by Kummer in 1835 in an attempt to prove that certain rings were unique factorization domains. [4, p. 157] In this section we will explore the interconnection between the ideals of a ring, and the prime and irreducible elements of the ring.

We begin by recalling that many ideals can be generated with only one element. In fact, many rings, such as the integers $\mathbb{Z}$, are such that every ideal is generated by only one element. We called such rings *principal ideal rings,* or PIRs. When the ring is also a domain, we call it a *principal ideal domain,* or PID. In fact, PIDs are so common that it is somewhat tricky to find an example of a UFD that is not a PID.

Consider the ring $R = \mathbb{Z}[x, y]$. We saw by corollary 12.6 that this is a UFD. We would now like to show that this is *not* a PID. Consider the ideal of elements without a constant term. This ideal can be expressed as (x, y), but since both x and y are in this ideal, we cannot express this ideal as the multiples of some polynomial. Thus, it requires at least two elements to generate this ideal in $\mathbb{Z}[x, y]$. Thus, this ideal is not a principal ideal, so $\mathbb{Z}[x, y]$ is not a PID, even though it is a UFD.

DEFINITION 12.11 Let R be a commutative ring, and let P be a nontrivial ideal of R. (Thus, P is neither $\{0\}$ nor R.) We say that P is a *prime ideal* if, whenever x and y are in R, and $x \cdot y$ is in P, then either x or y is in P.

When we first defined a prime element of a ring, we were careful to mention that the ring did not have to be an integral domain. By defining prime elements for all commutative rings, we open the door to showing a connection between prime ideals and prime elements.

PROPOSITION 12.6

Let R be a commutative ring with an identity. Then p is a prime element of R if, and only if, the principal ideal (p) is a prime ideal.

PROOF Suppose that p is prime. Then p is neither 0 nor a unit, so (p) cannot be the zero ring. If $(p) = R$, then there must be some element of R that makes $p \cdot x = 1$. But this is impossible, since p is not a unit. Thus, (p) would be a nontrivial ideal of R. Now suppose that $x \cdot y$ is in (p). Then there must be some z such that $x \cdot y = p \cdot z$. Since p is prime, either x or y is a multiple of p. So either x or y is in (p), making (p) a prime ideal.

Now suppose that (p) is a prime ideal. Then (p) is neither $\{0\}$ nor R, so p is neither 0 nor a unit. If $x \cdot y$ is a multiple of p, then $x \cdot y$ would be in (p). Since (p) is a prime ideal, either x or y would then be in (p). But this would indicate that x or y is a multiple of p. Thus, p is a prime element of R. □

Although this proposition refers to principal ideals, it is certainly possible for an ideal to be a prime ideal without being even a principal ideal. For example, the ideal (x, y) of the ring $\mathbb{Z}[x, y]$ is not a principal ideal, yet it is a prime ideal. To see this, note that we can characterize the ideal as

$$(x, y) = \{f(x, y) \in \mathbb{Z}[x, y] \quad | \quad f(0, 0) = 0\}.$$

Thus, if $f(x, y) \cdot g(x, y)$ is in (x, y), we have $f(0, 0) \cdot g(0, 0) = 0$, so either $f(0, 0) = 0$ or $g(0, 0) = 0$. So (x, y) is a prime ideal.

Although proposition 12.6 gives us a test for determining whether an element is prime, to implement this we need a way to see whether an ideal is a prime ideal.

PROPOSITION 12.7

Let R be a commutative ring with identity, and let P be a nontrivial ideal of R. Then P is a prime ideal if, and only if, the quotient ring R/P has no zero divisors.

PROOF Assume that P is a prime ideal. Let us suppose that the product of two elements of R/P, $a + P$ and $b + P$, is the zero element. That is,

$$(a + P) \cdot (b + P) = a \cdot b + P = 0 + P.$$

This implies that $a \cdot b$ is in P. Since P is a prime ideal, either a or b is in P. Thus, either

$$a + P = 0 + P \qquad \text{or} \qquad b + P = 0 + P.$$

Thus, we have shown that R/P has no zero divisors.

Now suppose that R/P has no zero divisors. If $a \cdot b$ is in P, then we have the following holding in R/P:

$$(a + P) \cdot (b + P) = a \cdot b + P = 0 + P.$$

Since R/P has no zero divisors, either $a + P$ or $b + P$ must be equal to $0 + P$. Thus, either a or b is in P, and since P is a nontrivial ideal, P is a prime ideal. □

Let us try to use this proposition to find the prime elements of the following familiar commutative ring:

InitRing
Define[4a, 0]; Define[2b, 0]
Define[a.a, a]; Define[b.b, b]
Define[a.b, 0]; Define[b.a, 0]
R = Ring[{a, b}]

```
gap> InitRing("a","b");
gap> DefineRing("R",[4,2],[[a,0],[0,b]]);
```

We determined that the element $2a + b$ was irreducible in this ring. Let us determine whether $2a + b$ is prime by computing the quotient ring $R/(2a + b)$.

First, we find the principal ideal generated by $2a + b$:

S = Ideal[R,{2a + b}]

```
gap> S := Ideal(R,[2*a+b]);
<two-sided ideal in <ring with 2 generators>, (1 generators)>
gap> List(S);
[ 0*a, b, 2*a, 2*a+b ]
```

This forms a nontrivial ideal, so we can now consider the quotient ring.

Q = Coset[R, S]
$\{\{0, 2a, b, 2a + b\}, \{a, 3a, a + b, 3a + b\}\}$
Q[[2]].Q[[2]]
$\{a, 3a, a + b, 3a + b\}$

In GAP, we can either list the cosets, or we can have GAP create a isomorphic copy of the quotient ring through the first ring isomorphism theorem (10.2).

```
gap> f := NaturalHomomorphismByIdeal(R,S);
[ a, b ] -> [ q1, 0*q1 ]
gap> Q := Image(f,R);
<ring with 2 generators>
gap> MultTable(Q);

*   |0*q1  q1
----|----------
0*q1|0*q1  0*q1
q1  |0*q1  q1
```

The quotient ring has only two elements, and in fact is isomorphic to Z_2. So $2a + b$ is a prime element of R.

We are mainly interested in finding the prime elements of an *infinite* ring. *Mathematica* can still often help us out, since the quotient ring $R/(p)$ will usually be finite.

Consider the ring $\mathbb{Z}[\sqrt{-5}]$. We saw in the last section that 3 was an irreducible element. To see whether this is a prime element, we need to determine the ring $\mathbb{Z}[\sqrt{-5}]/(3)$. Since 3 is in the ideal (3), every element multiplied by 3 in the quotient ring must be 0. Thus, the characteristic of the quotient ring is 3. We can start by defining the quotient ring as a domain:

InitDomain[3]

If we denote the element $\sqrt{-5} + (3)$ by a, then $a^2 = -5 + (3)$. Thus, we can define

Define[a^2, –5]

We now can see the quotient ring as the ring generated by 1 and a:

R = Ring[{1, a}]

This ring has nine elements. However, the command

CheckField[{1, a}]

reveals that this quotient ring has zero divisors. Thus, 3 is not a prime element of $\mathbb{Z}[\sqrt{-5}]$. We can form this same ring in GAP, but we have to plan ahead to see that the quotient ring will have nine elements.

```
gap> InitRing("e","a");
gap> DefineRing("R",[3,3],[[e,a],[a,-5*e]]);
gap> NumberElements := true;
true
gap> MultTable(R);

*       |1  2  3  4  5  6  7  8  9
--------+------------------------
0*e     |1  1  1  1  1  1  1  1  1
a       |1  4  7  2  5  8  3  6  9
2*a     |1  7  4  3  9  6  2  8  5
e       |1  2  3  4  5  6  7  8  9
e+a     |1  5  9  5  9  1  9  1  5
e+2*a   |1  8  6  6  1  8  8  6  1
2*e     |1  3  2  7  9  8  4  6  5
2*e+a   |1  6  8  8  1  6  6  8  1
2*e+2*a |1  9  5  9  5  1  5  1  9
```

At this point you may be wondering whether there are *any* prime elements in the ring $\mathbb{Z}[\sqrt{-5}]$. Consider the element $3 + 2\sqrt{-5}$. Defining the ring $\mathbb{Z}[\sqrt{-5}]/(3 + 2\sqrt{-5})$ in *Mathematica* or GAP is a bit trickier since the characteristic must be an integer. But note that

$$(3 + 2\sqrt{-5}) \cdot (3 - 2\sqrt{-5}) = 29.$$

Thus, 29 is in the ideal $(3 + 2\sqrt{-5})$, which we can use for the characteristic.

To reduce the elements further, we would like to find an integer within the coset

$$a + (3 + 2a) = \sqrt{-5} + (3 + 2\sqrt{-5}).$$

After experimenting, we find that the integer 13 is in this coset. This experiment shows that the element $\sqrt{-5}$ is equivalent to 13 in the ring $\mathbb{Z}[\sqrt{-5}]/(3+2\sqrt{-5})$. Thus, every element in the ring is equivalent to an integer. The quotient ring will have 29 elements, which is obviously isomorphic to the field Z_{29}. Thus, we have found a prime element for this ring.

We have seen that proposition 12.7 is a useful way of determining whether an element is prime. Let us use this proposition to show that in a principal ideal domain, irreducible elements are also prime elements. This amounts to showing that $R/(p)$ has no zero divisors whenever p is irreducible. However, we can actually prove more, which will be very useful later on.

LEMMA 12.6

Let R be a principal ideal domain, and let p be an irreducible element of R. Then the quotient ring $R/(p)$ is a field.

PROOF Since R is an integral domain, it is clear that $R/(p)$ is a commutative ring, and contains the identity element $1 + (p)$. Thus, we have to show that all nonzero elements of $R/(p)$ have an inverse. Let $x + (p)$ be a nonzero element of $R/(p)$. We immediately have that x is not a multiple of p. Thus, we can consider the ideal generated by both x and p, that is, (x, p).

Since R is a PID, there is some element d in R such that $(x, p) = (d)$. Then both x and p would be multiples of d. But we already observed that x is not a multiple of p, so d cannot be a multiple of p. But p is irreducible, so d must be a unit. Then $(d) = R$, and so $(x, p) = R$. This means that there are elements u and v in R such that

$$x \cdot u + p \cdot v = 1.$$

We now claim that $u + (p)$ is our sought-after inverse. Note that

$$[x + (p)] \cdot [u + (p)] = x \cdot u + (p) = x \cdot u + p \cdot v + (p) = 1 + (p).$$

Since every nonzero element of $R/(p)$ is invertible, we have that $R/(p)$ is a field. □

From this lemma, it is easy to see that an irreducible element of a PID must also be a prime element. Thus, we are on our way to showing that a PID is a unique factorization domain. By proposition 12.5, we only need to show that every non-invertible element can be expressed as a product of irreducible factors. In order to eliminate the possibility of an "infinite chain" of irreducible elements, each one dividing the previous, we will use the following lemma.

LEMMA 12.7
Let R be a principal ideal ring. If there is an infinite sequence of larger and larger ideals of R satisfying

$$I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \subseteq I_n \subseteq I_{n+1} \subseteq \cdots,$$

then there exists an integer m such that $I_n = I_m$ for all $n > m$.

PROOF Since we have an infinite sequence of ideals, we can consider taking the union of all of them:

$$I = \bigcup_{n=1}^{\infty} I_n.$$

Let us show that I is an ideal of R. Note that any element of I is in I_k for some integer k. In fact, if x and y are two elements of I, we can pick the larger of the two values of k to show that x and y are both in I_k. Then $x \pm y$ is in I_k, since I_k is an ideal. Thus $x \pm y$ is in I. This shows that I is a subgroup of R under addition. Now let z be in R. Then $x \cdot z$ and $z \cdot x$ are both in I_k, so $x \cdot z$ and $z \cdot x$ are in I. Therefore, $I \cdot R = R \cdot I = I$. This shows that I is an ideal.

Since R is a principal ideal ring, there is some element a in R such that $I = (a)$. Then a is in I_m for some m. But I_m is contained in I, so we must have that $I = I_m$. Thus, $I_n = I_m$ for all $n > m$. □

We now have all we need to show that a PID is in fact a UFD.

THEOREM 12.4: The Principal Ideal Domain Theorem
Every principal ideal domain is a unique factorization domain.

PROOF Our strategy is to first show that an irreducible element is a prime element, and then show that every element is a finite product of irreducible elements. Let p be an irreducible element of R, which is a PID. By lemma 12.6 $R/(p)$ is a field, so it certainly has no zero divisors. Thus, by proposition 12.7, (p) is a prime ideal, so by proposition 12.6, p is prime. Let us now show that every non-invertible element of R can be written as a product of irreducible elements. Suppose this is not true for some element x_0. Then x_0 is not irreducible, so we can find elements x_1 and y_1 in R such that $x_1 \cdot y_1 = x_0$. But x_1 and y_1 cannot both be irreducible, so we can assume x_1 is reducible. By induction we can continue this process to form a sequence

$$\{x_0, x_1, x_2, x_3, \cdots\}$$

for which each term in the sequence divides the previous term. Then we have an infinite chain of ideals,

$$(x_0) \subseteq (x_1) \subseteq (x_2) \subseteq (x_3) \subseteq \cdots.$$

By lemma 12.7, there is a number m such that $(x_n) = (x_m)$ for all $n > m$. But this contradicts the fact that all of the x_n are reducible. Therefore every element of R can be expressed as a product of irreducible elements. By proposition 12.5, R is a unique factorization domain. □

This theorem reveals the most important use of principal ideal domains—it enables us to find unique factorization domains. For example, $\mathbb{Z}$ was proven to be a PID from proposition 10.3, so we now can see that $\mathbb{Z}$ is a UFD, a result that was promised in section 1.3.

It should be noted that not all unique factorization domains are PIDs—in fact we discovered that $\mathbb{Z}[x, y]$ is not a PID, even though it is a UFD. However, many of the important unique factorization domains are also principal ideal domains.

Of course, there still is the problem of how to determine whether an integral domain is a PID. In the next section, we will find the main way of determining whether a certain domain is in fact a PID, which would then prove that it is a UFD.

12.4 Euclidean Domains

We have already seen the importance of principal ideal domains to determine whether a ring is a unique factorization domain. However, we still have the problem of determining whether a given integral domain is a principal ideal domain. This can usually be done quite easily.

For example, to show that $F[x]$ is a PID for any field F, we examine what the ideals could be. If I is a nontrivial ideal of $F[x]$, we can find a nonzero element $f(x)$ in I with the lowest degree. If $g(x)$ is also in I, then by the division algorithm

$$g(x) = f(x) \cdot q(x) + r(x),$$

with the degree of $r(x)$ less then $f(x)$. But $r(x)$ would also be in I, and since $f(x)$ has least degree of all the nonzero elements in I, we must have $r(x) = 0$. Therefore all elements of I are multiples of $f(x)$, so $I = (f(x))$.

Rather than making this a formal proposition, we want to study this example, since we can prove that *many* different domains are PIDs the same way. There were two keys to the proof that $F[x]$ was a PID: the fact that every polynomial had a degree, and the division algorithm. Whenever we have an integral domain that has a property like a division algorithm, there is a good chance that we can use this division algorithm to prove that the ring is a PID. Let us formulate what we mean by a "division algorithm."

DEFINITION 12.12 *An integral domain R is called a* Euclidean domain

if there is a function $\mu(x)$ defined on the nonzero elements of R such that the following three properties hold:

1. *$\mu(x)$ is a non-negative integer for every nonzero x in R.*
2. *Whenever both x and y are nonzero, $\mu(x \cdot y) \geq \mu(x)$.*
3. *For and x and y in R, with y nonzero, there exist elements q and r in R such that*
$$x = q \cdot y + r,$$
where either $r = 0$ or $\mu(r) < \mu(y)$.

The function $\mu(x)$ is called the Euclidean valuation *on R.*

Let us first look at some examples of Euclidean domains. Since this definition was modeled after the ring $F[x]$, it is expected that $F[x]$ is a Euclidean domain. The function $\mu(f(x))$ would be the degree of the polynomial $f(x)$. Properties 1 and 2 come from the definition of the degree, and lemma 11.1. Property 3 we observed in the division algorithm theorem (12.1). Thus, $F[x]$ is a Euclidean domain whenever F is a field.

However, there are many other examples of Euclidean domains. Consider the set of integers, $\mathbb{Z}$. We can use the absolute value for the valuation: $\mu(x) = |x|$. Clearly properties 1 and 2 hold, and the third property comes from modular arithmetic. Thus, $\mathbb{Z}$ is also a Euclidean domain.

Whenever we have a Euclidean domain, we can prove that the domain is a PID, using the exact same argument as we did for $F[x]$.

THEOREM 12.5: The Euclidean Domain Theorem
Every Euclidean domain is a principal ideal domain.

PROOF Let R be a Euclidean domain, and let $\mu(x)$ be the valuation. If I is an ideal, we consider the set

$$P = \{\mu(x) \mid x \in I, x \neq 0\}.$$

The set P consists of non-negative integers, so there is a smallest number in P. Pick an element y in I so that $\mu(y)$ is the minimal number in P. Then for any other x in I, we have

$$x = y \cdot q + r$$

for some q and r in R, with $\mu(r) < \mu(y)$. Then r is in I, but if r were nonzero, then this would contradict the minimality of $\mu(y)$. Thus, $r = 0$, and so x is a multiple of y. Since this is true for all x in I, we see that $I = (y)$. Thus, every ideal of R is a principal ideal, so R is a PID. ∎

We started this section by showing that $F[x]$ is a principal ideal ring whenever F is a field, but let us formally make this a corollary of the Euclidean domain theorem.

COROLLARY 12.7
Let F be a field. Then the ring of polynomials $F[x]$ is a principal ideal domain.

PROOF We have already seen that $F[x]$ is a Euclidean domain whenever F is a field. By the Euclidean domain theorem (12.5), $F[x]$ is a PID. □

The only problem with this definition of the Euclidean domain is that it gives no help in determining what the valuation function $\mu(x)$ should be. In fact, there may be many possible valuation functions for a given integral domain. See problem 12.29 for an alternative definition of a Euclidean domain that does not involve a valuation function.

For the remainder of this chapter, we will consider an interesting class of integral domains, some of which are Euclidean domains, and some that are not. This class of domains will help us to see some general techniques for finding a valuation function for a domain.

DEFINITION 12.13 Let n be an integer that is not divisible by the square of any integer other than 1. Then the ring $\mathbb{Z}[\sqrt{n}]$ is called a *quadratic domain.*

We have already worked with some examples of quadratic domains. For example, we found two possible ways to order the ring $\mathbb{Z}[\sqrt{2}]$, using ring homomorphisms.

The quadratic domain $\mathbb{Z}[\sqrt{n}]$ will always have two automorphisms, the identity mapping, and the automorphism

$$f(x + y\sqrt{n}) = x - y\sqrt{n}.$$

We define the function N as the product of the two automorphisms:

$$N(x + y\sqrt{n}) = (x + y\sqrt{n}) \cdot (x - y\sqrt{n}) = x^2 - y^2 n.$$

Note that $N(a)$ will always an integer.

At first glance it may be difficult to see what the $N(a)$ has to do with the Euclidean domains. Our goal is to construct a valuation function from $N(a)$. We first need to verify some elementary properties of this function. In the process, we will notice that these properties are still valid if we extend $N(a)$ to be defined on $\mathbb{Q}[\sqrt{n}]$.

LEMMA 12.8
Let $\mathbb{Z}[\sqrt{n}]$ be a quadratic domain, and let $N(x+y\sqrt{n}) = x^2 - y^2 n$. Then for the rings $Z[\sqrt{n}]$ and $\mathbb{Q}[\sqrt{n}]$,

1. *$N(a) = 0$ if, and only if, $a = 0$.*
2. *$N(a \cdot b) = N(a) \cdot N(b)$.*
3. *$N(\pm 1) = 1$.*

PROOF

1. It is easy to see that $N(0) = 0$ by definition. If $N(x + y\sqrt{n}) = 0$, then

   $$(x + y\sqrt{n}) \cdot (x - y\sqrt{n}) = x^2 - y^2 n = 0.$$

   If y is nonzero, then we find that $\sqrt{n} = |\frac{x}{y}|$, which is ridiculous since n is not a perfect square, and so $\sqrt{n}$ is irrational. Thus, $y = 0$, and hence x is also 0. So $N(a) = 0$ if, and only if, $a = 0$.

2. A quick computation shows that if $a = x_1 + y_1\sqrt{n}$, and $b = x_2 + y_2\sqrt{n}$, then

   $$a \cdot b = \left(x_1 + y_1\sqrt{n}\right) \cdot \left(x_2 + y_2\sqrt{n}\right) = (x_1 \cdot x_2 + y_1 \cdot y_2 \cdot n) + (x_1 \cdot y_2 + y_1 \cdot x_2)\sqrt{n}.$$

   So

   $$\begin{aligned} N(a \cdot b) &= (x_1 \cdot x_2 + y_1 \cdot y_2 \cdot n)^2 - (x_1 \cdot y_2 + y_1 \cdot x_2)^2 \cdot n \\ &= x_1^2 x_2^2 + 2x_1 x_2 y_1 y_2 n + y_1^2 y_2^2 n^2 - x_1^2 y_2^2 n - 2x_1 x_2 y_1 y_2 n - y_1^2 x_2^2 n \\ &= x_1^2 x_2^2 + y_1^2 y_2^2 n^2 - x_1^2 y_2^2 n - y_1^2 x_2^2 n \\ &= (x_1^2 - y_1^2 n) \cdot (x_2^2 - y_2^2 n) = N(a) \cdot N(b). \end{aligned}$$

3. This is easy, since $\pm 1 = \pm 1 + 0\sqrt{n}$. So $N(\pm 1) = (\pm 1)^2 - 0 \cdot n = 1$. ∎

We can use the $N(a)$ function to prove that $\mathbb{Q}[\sqrt{n}]$ is a field.

COROLLARY 12.8
Let n be an integer that is not divisible by the square of any integer greater than 1. Then the ring $\mathbb{Q}[\sqrt{n}]$ is a field.

PROOF Since $\mathbb{Q}[\sqrt{n}]$ is obviously a commutative ring with an identity, all we need to show is that every nonzero element has an inverse. Let $b = x + y\sqrt{n}$ be a nonzero element. Then $N(b)$ is nonzero by lemma 12.8. Consider the element

$$c = (x - y\sqrt{n})/N(b).$$

Then

$$b \cdot c = (x + y\sqrt{n}) \cdot (x - y\sqrt{n})/N(b) = N(b)/N(b) = 1.$$

So every nonzero element has an inverse. Thus, $\mathbb{Q}[\sqrt{n}]$ is a field. □

Using these three properties of the function $N(a)$, we are able to determine at least some of the irreducible elements of the ring $\mathbb{Z}[\sqrt{n}]$.

PROPOSITION 12.8

Let $\mathbb{Z}[\sqrt{n}]$ be a quadratic domain, and let $N(x + y\sqrt{n}) = x^2 - y^2 n$. Then

1. *$N(a) = \pm 1$ if, and only if, a is a unit in $\mathbb{Z}[\sqrt{n}]$, and*
2. *If $N(a)$ is a prime number in $\mathbb{Z}$, then a is an irreducible element of $\mathbb{Z}[\sqrt{n}]$.*

PROOF Suppose that $N(a) = N(x + y\sqrt{n}) = \pm 1$. Consider the element

$$b = (x - y\sqrt{n})/N(a).$$

Then

$$a \cdot b = (x + y\sqrt{n}) \cdot (x - y\sqrt{n})/N(a) = N(a)/N(a) = 1.$$

So a has an inverse, and therefore is a unit in $\mathbb{Z}[\sqrt{n}]$.

Now suppose that a is a unit in $\mathbb{Z}[\sqrt{n}]$. Then a has an inverse, a^{-1}. Then

$$1 = N(1) = N(a \cdot a^{-1}) = N(a) \cdot N(a^{-1}),$$

which shows that $N(a)$ must be ± 1.

Now suppose that $N(a) = p$, a prime number in $\mathbb{Z}$, and that $a = b \cdot c$. Then

$$p = N(a) = N(b \cdot c) = N(b) \cdot N(c).$$

Since p is prime, either $N(b)$ or $N(c)$ is ± 1. So either b or c must be a unit in $\mathbb{Z}[\sqrt{n}]$, so a is irreducible in $\mathbb{Z}[\sqrt{n}]$. □

We can now use the Euclidean function $\mu(x) = |N(x)|$ to prove the following.

PROPOSITION 12.9

The integral domains $\mathbb{Z}[\sqrt{-2}]$, $\mathbb{Z}[\sqrt{-1}]$, $\mathbb{Z}[\sqrt{2}]$, and $\mathbb{Z}[\sqrt{3}]$ are Euclidean domains.

PROOF Let us work with all four domains at the same time by considering $\mathbb{Z}[\sqrt{n}]$, where $n = -2, -1, 2$, or 3.

If we let $\mu(x) = |N(x)|$, then clearly $\mu(x)$ is a non-negative integer. Furthermore, $\mu(x) = 0$ only when $x = 0$. Thus, if u and v are two elements of $\mathbb{Z}[\sqrt{n}]$, then

$$\mu(u \cdot v) = |N(u \cdot v)| = |N(u)| \cdot |N(v)| = \mu(u) \cdot \mu(v) \leq \mu(u) \cdot 1 = \mu(u).$$

So the first two conditions for the valuation function are easily satisfied. The last condition is harder to prove. We need to show that for any x and y in $\mathbb{Z}[\sqrt{n}]$, with y nonzero, there are elements q and r such that

$$x = q \cdot y + r,$$

with either $r = 0$, or $\mu(r) < \mu(y)$. We can consider x and y to be in $\mathbb{Q}[\sqrt{n}]$, which is a field from corollary 12.8, so we can compute

$$t = x \cdot y^{-1} = u + v\sqrt{n}.$$

Of course, t will be in $\mathbb{Q}[\sqrt{n}]$ instead of $\mathbb{Z}[\sqrt{n}]$, so we cannot use this for our q. However, we can find an element "closest" to t in $\mathbb{Z}[\sqrt{n}]$ by finding the integers p and k nearest to u and v. That is, we will select integers p and k such that

$$(*) \qquad |p - u| \leq \frac{1}{2} \qquad \text{and} \qquad |k - v| \leq \frac{1}{2}.$$

We now let $q = p + k\sqrt{n}$, which is in $\mathbb{Z}[\sqrt{n}]$. The remainder r would be given by $q \cdot y - x$. All we need to do is show that $r = 0$, or $\mu(r) < \mu(y)$.

Now, the norm $N(x)$ is valid on $\mathbb{Q}[\sqrt{n}]$, so we can compute

$$N(q - t) = N\left((p - u) + (k - v)\sqrt{n}\right) = (p - u)^2 - n(k - v)^2.$$

By $(*)$ we see that if $n > 0$,

$$-n/4 \leq (p - u)^2 - n(k - v)^2 \leq 1/4.$$

On the other hand, if $n < 0$, then

$$0 \leq (p - u)^2 - n(k - v)^2 \leq (1 - n)/4.$$

Thus, as long as $-2 \leq n \leq 3$ we have that

$$|N(q - t)| = |(p - u)^2 - n(k - v)^2| \leq 3/4 < 1.$$

Thus,

$$\begin{aligned} \mu(r) &= |N(r)| = |N(q \cdot y - x)| \\ &= |N((q - x \cdot y^{-1}) \cdot y)| \\ &= |N(q - t)| \cdot |N(y)| \\ &< |N(y)| = \mu(y). \end{aligned}$$

Therefore, the function $\mu(x)$ serves as a valuation function on $\mathbb{Z}[\sqrt{n}]$, and so $\mathbb{Z}[\sqrt{n}]$ is a Euclidean domain for $n = -2, -1, 2$, or 3. □

One of these four domains has special applications. The ring $\mathbb{Z}[\sqrt{-1}] = \mathbb{Z}[i]$ is called the domain of *Gaussian integers*. *Mathematica*'s **FactorInteger** command can find the prime factorization over the Gaussian integers by setting a condition "GaussianIntegers" to true. For example, we can factor the number 5 as follows:

FactorInteger[5, GaussianIntegers –> True]

The GAP `Factor` command allows one to put the ring as the second argument.

```
gap> Factor(5, GaussianIntegers);
[ 2-E(4), 2+E(4) ]
```

This reveals that $5 = (2 - i) \cdot (2 + i)$. By investigating further the divisibility properties of $\mathbb{Z}[i]$, one can prove the classic "two squares theorem" of Fermat: Every prime number of the form $4n + 1$ is the sum of two squares. (See problem 13.18.) It is interesting that the study of domains other than the familiar integers yields new information about the integers.

Since every Euclidean domain is a PID, the natural question to ask is whether there is a PID which is *not* a Euclidean domain. There actually are such domains, although known examples are rare. The simplest example is $\mathbb{Z}[(1 + \sqrt{-19})/2]$, but it is tricky to prove that this example works for two reasons. First of all, to show that this ring is *not* a Euclidean domain, we must show that no valuation function $\mu(x)$ can be defined whatsoever. Problem 12.29 gives an alternative way to define a Euclidean domain that does not depend on a valuation function, and hence helps in showing that $\mathbb{Z}[(1+\sqrt{-19})/2]$ is not a Euclidean domain. But then we must show that this ring is still a PID, which is especially hard since the main tool for proving that a domain is a PID is the Euclidean domain theorem (12.5). For a sketch of how this is proven, see problems 12.46 to 12.53. A similar proof can be used to show that $\mathbb{Z}[(1 + \sqrt{-43})/2]$, $\mathbb{Z}[(1 + \sqrt{-67})/2]$, and $\mathbb{Z}[(1 + \sqrt{-163})/2]$ are PIDs, but not Euclidean domains.

Problems for Chapter 12

Interactive Problems

12.1 Use the *Mathematica* command **InterpolatingPolynomial** or GAP's `InterpolatedPolynomial` to find a third degree polynomial such that $f(n) = n!$ for $n = 1, 2, 3$, and 4. How close is $f(5)$ to 120?

12.2 Use GAP or *Mathematica* to determine whether $x^3 + 2x^2 + 3x + 2$ is irreducible over Z_5.

12.3 Use GAP or *Mathematica* to determine whether $x^3 + 2x^2 + 3x + 5$ is irreducible over Z_7.

12.4 Define the domain $\mathbb{Z}[\sqrt{6}]$ in *Mathematica* as follows:

InitDomain[0]
Define[a^2, 6]

Show that the element $u = 5 + 2a$ is a unit by finding its inverse. Use the element u to find yet another unit of $\mathbb{Z}[\sqrt{6}]$.

12.5 Use *Mathematica* to show that the ring $\mathbb{Z}[\sqrt{6}]/(11)$ has no zero divisors. Use this to prove that 11 is a prime element of $\mathbb{Z}[\sqrt{6}]$.

12.6 Use the *Mathematica* command

FactorInteger[2, GaussianIntegers –> True]

or the GAP command

```
gap> Factor(2, GaussianIntegers);
```

to determine whether 2 is prime in the domain $\mathbb{Z}[i]$. Try this using the numbers 3, 5, 7, 11, 13, 17, 19, 23, 29, and 31 in place of 2. Which of these numbers are prime in the domain $\mathbb{Z}[i]$?

Non-Interactive Problems

12.7 Use the division algorithm to determine polynomials $q(x)$ and $r(x)$ in $\mathbb{Q}[x]$ such that

$$2x^3 + 3x^2 - 5x + 4 = (2x^2 - x + 1) \cdot q(x) + r(x),$$

where $r(x)$ has degree less than 2.

12.8 Use the division algorithm to determine polynomials $q(x)$ and $r(x)$ in $Z_2[x]$ such that

$$x^5 + x^3 + x^2 + x = (x^3 + x^2 + 1) \cdot q(x) + r(x),$$

where $r(x)$ has degree less than 3.

12.9 Find a quadratic polynomial $f(x)$ such that $f(-1) = 6$, $f(1) = 2$, and $f(2) = 9$.

Hint: Either solve three equations for three unknowns, or use the proof of corollary 12.3.

12.10 Find a quadratic polynomial in $Z_3[x]$ such that $f(0) = f(1) = 2$, and $f(2) = 0$.

12.11 Prove that $x^2 + 5$ is irreducible over the field $\mathbb{R}$ of real numbers.

12.12 Prove that $x^3 - 3x + 3$ is irreducible over the field $\mathbb{Q}$ of rational numbers.

Hint: Prove that it is irreducible over the integers, and use lemma 12.4.

12.13 Show that $x^3 - 9$ is irreducible over the field Z_{13}.

12.14 Find the factorization of $x^3 + 2x^2 + 2$ over the field Z_3.

12.15 Find the factorization of $x^3 + 2x^2 + 2$ over the field Z_5.

12.16 Find the factorization of $x^3 + 2x^2 + 2$ over the field Z_7.

12.17 Find the factorization of $x^4 + 2x^2 + 2$ over the field Z_5.

12.18 Let F be a field that is contained in a larger field K. Let $f(x)$ and $g(x)$ be two polynomials in $F[x]$ that are coprime in $F[x]$. Show that $f(x)$ and $g(x)$ are also coprime in $K[x]$.

12.19 Show that $\mathbb{Z}[\sqrt{-5}]$ is not a principal ideal domain by finding an ideal of this ring that is not a principal ideal.

Hint: Consider the ideal $(2, 1 + \sqrt{-5})$.

12.20 Find all of the irreducible elements of Z_{12}.

Hint: First find all of the units. Construct a multiplication table of the non-units. Which elements do not appear in the interior of the table?

12.21 Find all of the prime ideals of Z_{12}. (Note that this ring has zero divisors.)

12.22 Find all of the prime elements of Z_{12}. (Note that this ring has zero divisors. See problem 12.21.)

12.23 Find all of the irreducible elements of Z_{18}. (See the hint for problem 12.20.)

12.24 Find all of the prime ideals of Z_{18}.

12.25 Find all of the prime elements of Z_{18}. (See problem 12.24.)

12.26 Show that the ring Z_8 has unique factorization, even though it is not an integral domain.

12.27 Can a field have irreducible or prime elements? Explain.

12.28 Let R be an integral domain. Prove that if every nontrivial ideal of R is a prime ideal, then R is a field.

Hint: If x is an element of R, show that x is contained in x^2R.

12.29 Suppose that R is an integral domain. Let S_0 be the set containing all units of R, along with the zero element. Let S_1 be the set of all elements x such that either $x = 0$ or

$$(x) + S_0 = R.$$

(That is, every element of R can be written as a multiple of x plus an element of S_0.) Define S_i inductively as the set of elements x such that either $x = 0$ or

$$(x) + S_{i-1} = R.$$

Prove that R is a Euclidean domain if, and only if, every element of R is in S_n for some n.

Hint: Let $\mu(x)$ be the smallest value of n for which x is in S_n.

12.30 Let R be a commutative ring, and let I be an ideal of R. If P is a prime ideal of I, prove that P is an ideal of R.

12.31 Let R be a PID. Prove that every element that is neither 0 nor a unit is divisible by some prime element.

12.32 Show that the elements q and r in part 3 of the definition of a Euclidean domain are not necessarily unique.

Hint: In $\mathbb{Z}[i]$, let $x = -4 + i$, $y = 5 + 3i$. Consider $q = -1 + i$ and $q = -1$.

12.33 Consider the subring of the elements of $\mathbb{Q}[x]$ for which the constant term is an integer. Show that this subring is not a UFD.

Hint: Show that the only units are ± 1, and that 2 is irreducible. Consider the sequence $x, x/2, x/4, x/8, \ldots x/(2^n), \ldots$.

12.34 Let D be a Euclidean domain, and let μ be the valuation function. Show that u is a unit in D if, and only if, $\mu(u) = \mu(1)$.

12.35 Let D be a Euclidean domain, and let μ be the valuation function. Show that if a and b are associates, then $\mu(a) = \mu(b)$.

12.36 Show that $\mathbb{Z}[\sqrt{-6}]$ is not a unique factorization domain.

Hint: Factor 10 in two ways.

12.37 Prove that 7 is prime in $\mathbb{Z}[\sqrt{6}]$.

Hint: First show that $x^2 - 6y^2 \equiv 0 \pmod 7$ only when x and y are both $0 \pmod 7$.

12.38 Show that if $n \equiv 3 \pmod 4$, then n cannot be expressed as the sum of two square integers.

12.39 If $a^2 + b^2$ is a prime number in the ordinary sense, prove that $a + bi$ is a prime number in the domain $\mathbb{Z}[i]$.

Hint: Use proposition 12.8.

12.40 If $p = a^2 + b^2$ is a prime number in the ordinary sense, find the prime factorization of p in the domain $\mathbb{Z}[i]$. (See problem 12.39.)

12.41 Let $p > 0$ be a prime number in the ordinary sense. Show that p factors in the larger domain $\mathbb{Z}[i]$ if, and only if, there are two integers a and b for which $p = a^2 + b^2$. (See problem 12.40.)

12.42 Suppose that n is an integer for which $\sqrt{4n+1}$ is irrational. Let

$$q = \frac{1 + \sqrt{4n+1}}{2},$$

and consider the domain $\mathbb{Z}[q] = \{x + yq \mid x, y \in \mathbb{Z}\}$. Define the function $N(a)$ on $\mathbb{Z}[q]$ by

$$\begin{aligned} N(x + yq) &= \left(x + y\left(\frac{1 + \sqrt{4n+1}}{2}\right)\right) \cdot \left(x + y\left(\frac{1 - \sqrt{4n+1}}{2}\right)\right) \\ &= x^2 + xy - ny^2. \end{aligned}$$

Show that $N(x)$ satisfies the properties of lemma 12.8, that is, $N(a) = 0$ if, and only if, $a = 0$, $N(a \cdot b) = N(a) \cdot N(b)$, and $N(\pm 1) = 1$. These domains are called *semi-quadratic domains.*

12.43 Prove proposition 12.8 for the semi-quadratic domains $\mathbb{Z}[q]$ of problem 12.42.

12.44 Show that $\mathbb{Z}[(1 + \sqrt{-3})/2]$ is a Euclidean domain. This is the ring of *Eulerian integers.* (See problems 12.42 and 12.43.)

Hint: Use the same trick used in proposition 12.9. Since $\mathbb{Q}[q] = \mathbb{Q}[\sqrt{-3}]$ is a field by corollary 12.8, we can find $t = x \cdot y^{-1} = u + vq$ in $\mathbb{Q}[q]$, and then round u and v to the nearest integer to find an element in $\mathbb{Z}[q]$.

12.45 Show that $\mathbb{Z}[(1 + \sqrt{5})/2]$ is a Euclidean domain. This ring is called the *Golden ratio domain.* (See the hint for problem 12.44.)

12.46 Show that the only units of $\mathbb{Z}[(1 + \sqrt{-19})/2]$ are ± 1.

Hint: Use problems 12.42 and 12.43 with $n = -5$.

12.47 Show that 2 and 3 are prime numbers in $\mathbb{Z}[(1 + \sqrt{-19})/2]$.

Hint: Use problems 12.42 and 12.43. When can $x^2 + xy + 5y^2$ be even or a multiple of 3?

12.48 Use problem 12.29 to show that $\mathbb{Z}[(1+\sqrt{-19})/2]$ is *not* a Euclidean domain.

Hint: Use problems 12.46 and 12.47 to show that $S_1 = S_0$, and hence $S_i = S_0$ for all i.

12.49 For every complex number z, show that there is a $x \in \mathbb{Z}[(1+\sqrt{-19})/2]$ such that $|\mathrm{Re}(z-x)| \le 1/2$ and $0 \le \mathrm{Im}(z-x) \le \sqrt{19}/2$.

Hint: First find an x for which $0 \le \mathrm{Im}(z-x) \le \sqrt{19}/2$, then add an integer to x to get $|\mathrm{Re}(z-x)| \le 1/2$.

12.50 For every complex number z, show that there is a $y \in \mathbb{Z}[(1+\sqrt{-19})/2]$ such that either $|z-y| < 1$ or $|2z-y| < 1$.

Hint: First pick a y using problem 12.49, and draw a picture in the complex plane to show where y could be. Show that three circles of radius 1 centered at $(1\pm\sqrt{-19})/2$ and 0, and two circles of radius $1/2$ centered at $(1\pm\sqrt{-19})/4$ cover this region.

12.51 Let I be an ideal of $R = \mathbb{Z}[(1+\sqrt{-19})/2]$, and let m be a nonzero element of I for which $N(m)$ is as small as possible. (See problems 12.42 and 12.43 for the definition of $N(m)$.) Show that if $x \in I$, then there is a $y \in R$ such that $2x = my$.

Hint: Let $z = m^{-1}x \in \mathbb{Q}[\sqrt{-19}]$. We can extend the $N(x)$ function to $\mathbb{Q}[\sqrt{-19}]$, so problem 12.50 shows that there is a $y \in R$ for which $N(m^{-1}x - y) < 1$ or $N(2m^{-1}x - y) < 1$.

12.52 Let I be an ideal of $R = \mathbb{Z}[(1+\sqrt{-19})/2]$, and let $m \in I$, $m \neq 0$ have minimum $N(m)$ as in problem 12.51. Show that if $x \in I$, but $x \notin (m)$, then m is a multiple of 2, and that $x = (m/2)y$ for some $y \in R$ that is not a multiple of 2.

Hint: Problem 12.47 shows that 2 is prime in R.

12.53 Show that $\mathbb{Z}[(1+\sqrt{-19})/2]$ is a PID.

Hint: Use problem 12.52 to show that if I is an ideal that is not a principal ideal, and m is the element of I with the least nonzero $N(m)$, then $(m/2)y\overline{y} \in I$, and hence $m/2 \in I$, but $N(m/2) < N(m)$.

Chapter 13

Finite Division Rings

13.1 Entering Finite Fields in *Mathematica*® or GAP

In this section we will experiment with finite fields using GAP and *Mathematica*. Although we have seen how integral domains can be entered into GAP and *Mathematica*, fields have additional properties that allow for shortcuts in this process.

We have already seen several examples of finite fields. The first example was the discovery that whenever p is prime, the ring Z_p forms a field with p elements. In chapter 3 we found another example of a finite field—the "complex numbers modulo 3." This ring was defined in *Mathematica* with the commands

InitDomain[3]
Define[i^2, −1]
K = Ring[{i}]

or in GAP by

```
gap> InitRing("e","i");
gap> DefineRing("K",[3,3],[[e,i],[i,-e]]);
```

Let's show that there is a connection between this field and the polynomials in Z_3. Since Z_3 is a subfield of K that was previously entered, we can enter the subfield simply as

Z3 = {0, 1, 2}

```
gap> Z3 := [0*e, e, 2*e];
[ 0*e, e, 2*e ]
```

We can also factor polynomials in the subfield $Z_3[x]$. In GAP, we will define the variable x to be over K.

```
gap> x := Indeterminate(K,"x");
x
gap> Factor(x^3 + x^2 + e, Z3);
[ x-e, x^2-x-e ]
```

Factor[x^3 + x^2 + 1, Z3]
$(2+x)\left(2+2x+x^2\right)$

Notice in particular that the polynomial x^2+1 is irreducible in $Z_3[x]$.

```
gap> Factor(x^2 + e,Z3);
[ x^2+e ]
```

Each element of the field K can be thought of as evaluating some polynomial in $Z_3[x]$ at $x = i$. Even though i is not an element of Z_3, we can consider any polynomial in $Z_3[x]$ as being also a polynomial in $K[x]$. This suggests that we should use the evaluation homomorphism

$$\phi_i : K[x] \to K.$$

However, we can restrict this homomorphism to apply only to polynomials in $Z_3[x]$.

$$\phi_i' : Z_3[x] \to K.$$

The image will still be all of K, since $\phi_i(x) = i$. The kernel of this homomorphism will consist of all polynomials in $Z_3[x]$ that yield 0 when evaluated at $x = i$. For example, x^2+1 is in the kernel, as are all multiples of x^2+1. In fact, if $f(x)$ is an element of the kernel, then $\text{GCD}(f(x), x^2+1)$ must be in the kernel, and x^2+1 is irreducible in $Z_3[x]$. Thus, the kernel must be precisely the multiples of x^2+1. This ideal can be described as (x^2+1), the ideal generated by x^2+1.

By the first ring isomorphism theorem (10.2), we now have that

$$K \approx Z_3[x]/(x^2+1)$$

since the field K is the image of the homomorphism ϕ_i'.

We can try a similar process to produce other fields. Recall that we tried to form a field by extending Z_5 by an element i, where $i^2 = -1$. However, we failed to produce a field, since the ring had zero divisors. We succeeded in producing the ring

$$K \approx Z_5[x]/(x^2+1)$$

but x^2+1 factors in Z_5: $(x+2)(x+3)$. This factorization apparently causes the zero divisors to appear in the quotient ring. Perhaps we should try using a polynomial that is irreducible in Z_5. We first define Z_5 in GAP or *Mathematica*:

InitDomain[5]
Z5 = Ring[{1}]

```
gap> InitRing("e");
gap> DefineRing("Z5",[5],[[e]]);
```

Next, we find a polynomial that is irreducible in Z_5.

Factor[x^2 + 2 x + 3, Z5]

```
gap> x := Indeterminate(Z5,"x");
x
gap> Factor(x^2 + 2*e*x + 3*e,Z5);
[ x^2+2*e*x+3*e ]
```

So x^2+2x+3 is irreducible over Z_5. To find a new field for which x^2+2x+3 has a zero, we will denote one of the zeros by the letter w. Then it is clear that $w^2 = -2w - 3$, so we can enter this into *Mathematica*.

Define[w^2, −2 w − 3]

Mathematica can now generate the ring containing w.

H = Ring[{w}]

In gap, we have to define the ring from scratch.

```
gap> InitRing("e","w");
gap> DefineRing("H",[5,5],[[e,w],[w,-2*w-3*e]]);
gap> Size(H);
25
```

Although the ring formed has 25 elements, we can have the *Mathematica* command

CheckField[{1, w}]

verify that this is indeed a field. In GAP, we can list the inverses of all of the elements.

```
gap> List(H, x -> 1/x);
[ fail, e+2*w, 3*e+4*w, 2*e+w, 4*e+2*w, e, 2*e+2*w, 3*e+2*w, w,
  3*e+w, 3*e, 3*w, e+w, 4*e+3*w, 4*e+w, 2*e, e+4*w, e+2*w,
  4*e+4*w, 2*w, 4*e, 2*e+4*w, 4*w, 2*e+3*w, 3*e+3*w ]
```

Since only one element fails to have an inverse (namely `0*e`), this is a field. As in the case of $Z_3[x]/(x^2+1)$, we can describe this field as

$$Z_5[x]/(x^2+2x+3).$$

Thus we have found a way to form fields out of polynomial rings.

PROPOSITION 13.1

Let K be a field, and let $f(x)$ be an irreducible polynomial of $K[x]$. Then $K[x]/(f(x))$ is a field that contains K as a subfield.

PROOF Since K is a field, by corollary 12.7 $K[x]$ is a principal ideal domain. Since $f(x)$ is an irreducible element of $K[x]$, we have by lemma 12.6 that the quotient $H = K[x]/(f(x))$ is a field.

Finally, we need to show that the field H contains K as a subfield. Consider the mapping $f : K \to H$ given by

$$f(y) = y + (f(x)).$$

This is certainly a homomorphism, since it is a restriction of the natural homomorphism from $K[x]$ to $K[x]/(f(x))$. The kernel of f is just 0, so the image is isomorphic to K. Thus, $K[x]/(f(x))$ contains K as a subfield. □

DEFINITION 13.1 The field formed in proposition 13.1 is called the *extension field of K through the irreducible polynomial $f(x)$.*

The first step is to determine the size of this new field.

PROPOSITION 13.2
Let p be a prime number, and let $A(x)$ be an irreducible polynomial in $Z_p[x]$ of degree d. Then the field $Z_p[x]/(A(x))$ has order p^d.

PROOF By the division algorithm theorem (12.1), every element $f(x)$ of $Z_p[x]$ can be written

$$f(x) = q(x) \cdot A(x) + r(x),$$

where either $r(x)$ is 0, or the degree of $r(x)$ is less than d. Thus, the typical element of K,

$$f(x) + (A(x)),$$

could be written as $r(x) + (A(x))$. Furthermore, the $r(x)$ is uniquely determined from the division algorithm theorem. Thus, there are as many elements in K as there are polynomials in $Z_p[x]$ with degree less than d, counting the zero polynomial. All such polynomials can be written

$$a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_{d-1}x^{d-1},$$

with each a_i between 0 and $p-1$, inclusively. Since there are d coefficients, each of which can be p different numbers, there are exactly p^d possible polynomials of degree less than d. Thus, $|K| = p^d$. □

Whenever a finite field is defined by an extension through an irreducible polynomial, the order of the field will be a power of a prime. We would like to show that all finite fields are produced in this way. So naturally we begin by showing that all finite fields have an order that is a power of a prime number.

PROPOSITION 13.3
Suppose K is a finite division ring. Then $|K| = p^n$ for some prime p and some integer n.

PROOF Let q be the order of K. From the additive structure of the ring, we see that $q \cdot x = 0$ for all x in K. Thus, the characteristic is positive, and by proposition 11.2, the characteristic is a prime number, p.

Suppose that q has a prime factor r other than p. Then the additive group of K must have a subgroup of order r, according to lemma 6.2. Hence $r \cdot x = 0$ for some element x in K. But this contradicts proposition 11.2, since r is not divisible by p. Therefore, q has no prime factors other than p, so $q = p^n$ for some integer n. □

According to this proposition, it is impossible to find a field of order 6. However, it is still possible to find a field of order 4. An irreducible polynomial of degree 2 in $Z_2[x]$ is $x^2 + x + 1$. Thus the commands

InitDomain[2]
Define[a^2, −a − 1]
F = Ring[{a}]

TABLE 13.1: Field of order 4

+	0	1	a	1+a
0	0	1	a	1+a
1	1	0	1+a	a
a	a	1+a	0	1
1+a	1+a	a	1	0

·	0	1	a	1+a
0	0	0	0	0
1	0	1	a	1+a
a	0	a	1+a	1
1+a	0	1+a	1	a

find a field of order 4 shown in table 13.1. The multiplication tables can be found in GAP.

```
gap> InitRing("e","a");
gap> DefineRing("F",[2,2],[[e,a],[a,-a-e]]);
gap> ResetTableOptions();
gap> AddTable(F);

+  |0*e  a    e    e+a
---+------------------
0*e|0*e  a    e    e+a
a  |a    0*e  e+a  e
e  |e    e+a  0*e  a
e+a|e+a  e    a    0*e

gap> MultTable(F);

*  |0*e  a    e    e+a
---+------------------
0*e|0*e  0*e  0*e  0*e
a  |0*e  e+a  a    e
e  |0*e  a    e    e+a
e+a|0*e  e    e+a  a
```

As we see from this example, it is fairly easy to enter finite groups into *Mathematica* or GAP, as long as they can be expressed as an extension field of Z_p through some irreducible polynomial of $Z_p[x]$. In the next section, we will show that all finite fields can be obtained in this way. In fact, our goal will be to classify *all* finite fields, which will give us a more natural way of defining the fields in GAP.

13.2 Properties of Finite Fields

In the last example we starting looking at examples of finite fields. In this section we want to explore the properties that all finite fields have in common.

We begin by observing that if F is a finite field, that the multiplicative group F^* must be a finite abelian group. If the field is of order p^n, the group F^* has order $p^n - 1$. For example, the field of order 4 has a multiplicative group of order 3, so this group must be isomorphic to Z_3. By studying the other fields that we created in the previous section, we discover that the multiplicative groups have one feature in common.

PROPOSITION 13.4
If F is a finite field, then the multiplicative group F^* is a cyclic group.

PROOF F^* is abelian, and so by the fundamental theorem of abelian groups (6.2),

$$F^* \approx Z_{d_1} \times Z_{d_2} \times Z_{d_3} \times \cdots \times Z_{d_n},$$

where the d_i are all powers of prime numbers. Let d be the least common multiple of the set $\{d_1, d_2, d_3, \ldots, d_n\}$. Then for all x in F^*, we have that $x^d = 1$. Thus, the polynomial $x^d - 1$ has $|F^*|$ solutions. By corollary 12.2, d must be at least $|F^*|$. But we also have

$$|F^*| = d_1 \cdot d_2 \cdot d_2 \cdots d_n,$$

so d is at most $|F^*|$. Thus, $d = |F^*|$, and so $d_1, d_2, d_3, \cdots, d_n$ are coprime. Therefore, the group F^* is cyclic. ∎

Now that the multiplicative group is completely understood for a finite field, let us turn our attention to the group of automorphisms on the field. We have previously seen examples where the group of automorphisms gave us insight into the structure of a ring, and finite fields are no exception. We begin by proving some basic lemmas in number theory.

LEMMA 13.1
If p is a prime, then

$$n^p \equiv n \pmod{p}$$

for all integers n.

PROOF Since Z_p^* is of order $p-1$, we have by corollary 3.2 that

$$n^{p-1} = 1$$

for all elements n in Z_p^*. (This result is commonly called Fermat's little theorem.) If we multiply both sides by n,

$$n^p = n,$$

we have a statement that is true for $n = 0$ as well. Thus, $n^p = n$ for all n in the ring Z_p. This statement, when converted into modular notation, becomes

$$n^p \equiv n \pmod{p}.$$ □

LEMMA 13.2
If F is a field of characteristic p, then for all $g \in F$, the polynomial

$$f(x) = (x+g)^p - x^p - g^p$$

is the zero polynomial in $F[x]$.

PROOF If $g = 0$, $f(x) = x^p - x^p = 0$, so the result is trivial. Let us suppose that g is nonzero.

Note that the leading term of $(x+g)^p$ is x^p, which will cancel in $f(x)$. Thus, $f(x)$ has degree at most $p-1$. Yet for every n, $n \cdot g$ is a root. Observe that

$$f(n \cdot g) = (n \cdot g + g)^p - (n \cdot g)^p - g^p = ((n+1)^p - n^p - 1) \cdot g^p.$$

By lemma 13.1,

$$(n+1)^p \equiv (n+1) \pmod{p}$$

and

$$n^p \equiv n \pmod{p}.$$

Thus,

$$(n+1)^p - n^p - 1 \equiv (n+1) - n - 1 \equiv 0 \pmod{p}.$$

So because F has characteristic p, we have $f(n \cdot g) = 0$. Since g is nonzero, the values

$$\{0, g, 2g, 3g, \cdots, (p-1)g\}$$

are all distinct in F. Thus, $f(x)$ has p distinct roots. But corollary 12.2 shows us that if $f(x)$ were nonzero, there would be at most $p-1$ roots. Thus, $f(x)$ must be the zero polynomial. □

We are now ready to produce one automorphism on a finite field, which we will use to generate all other automorphisms.

THEOREM 13.1: The Frobenius Automorphism Theorem
If F is a finite field of characteristic p, then the mapping

$$f : x \to x^p$$

forms an automorphism of F to itself. Furthermore, $f(y) = y$ if, and only if, y is in the subfield Z_p. This automorphism is called the Frobenius automorphism on F.

PROOF We first need to show that f is a homomorphism. If F is a field of characteristic p, then by lemma 13.2 we have that

$$(x+g)^p - x^p - g^p = 0$$

for all g in F. Thus, we have the identity

$$f(x+y) = (x+y)^p = x^p + y^p = f(x) + f(y).$$

It is also obvious that

$$f(x \cdot y) = (x \cdot y)^p = x^p \cdot y^p = f(x) \cdot f(y).$$

So f is a homomorphism. The kernel of f is obviously just 0, since $x^p = 0$ implies that $x = 0$, since F has no zero divisors. Therefore, the mapping is one-to-one. Since F is a finite field, we can use the pigeonhole principle to show that the mapping is also onto. Therefore, f is an automorphism.

Finally, we need to show that $f(y) = y$ if, and only if, y is in the subfield Z_p. Note that this subfield is generated by the multiplicative identity, 1:

$$Z_p = \{0, 1, 2, 3, \cdots, p-1\}.$$

By lemma 13.1, for any element in this subfield, $f(x) = x^p = x$. On the other hand, by corollary 12.2, the polynomial $x^p - x$ in $F[x]$ cannot have more than p roots in F. We have already found p solutions, so there cannot be anymore. Therefore, $f(y) = y$ if, and only if, y is in Z_p. □

Once we have one automorphism $f(x)$, we can consider creating other automorphisms such as $f(f(x))$ and $f(f(f(x)))$. It is not hard to determine the order of $f(x)$.

COROLLARY 13.1
Let F be a finite field of order p^n. Then the Frobenius automorphism is of order n in the group of automorphisms.

PROOF Note that the multiplicative group F^* has order $p^n - 1$. Thus, by corollary 3.2, for every element x in F^*, we have

$$x^{(p^n-1)} = 1.$$

Multiplying both sides by x gives us $x^{p^n} = x$ for all x in F^*, and also $x = 0$. Thus, this statement is true for all x in F.

We now note that

$$f^n(x) = \underbrace{f(f(f(\cdots(f(x))\cdots)))}_{n \text{ times}} = x^{p^n} = x.$$

for all x in F, so f^n yields the identity automorphism.

To show that the order of f is not less than n, suppose that the order was $d < n$. Then $f^d(x) = x^{p^d}$ would be x for all x. But then the polynomial

$$x^{p^d} - x$$

would have p^n solutions. This contradicts corollary 12.2, since $n > d$. Therefore, the order of the Frobenius automorphism is n. □

We next need to show a simple lemma to indicate how to apply the Frobenius automorphism to the set of polynomials over the field.

LEMMA 13.3
Any isomorphism f that maps an integral domain K to an integral domain M extends to an isomorphism mapping $K[x]$ to $M[x]$, with $f(x) = x$.

PROOF Suppose $f(x)$ is an isomorphism mapping K to M. If $w(x)$ is in $K[x]$, with coefficients a_i, we can define $f(w(x))$ by

$$f(w(x)) = f\left(\sum_{i=0}^{\infty} a_i x^i\right) = \sum_{i=0}^{\infty} f(a_i)x^i.$$

If $v(x)$ is another polynomial in $K[x]$ with coefficients b_i, then

$$\begin{aligned} f(w(x) + v(x)) &= f\left(\sum_{i=0}^{\infty}(a_i + b_i)x^i\right) = \sum_{i=0}^{\infty} f(a_i + b_i)x^i. \\ &= \sum_{i=0}^{\infty} f(a_i)x^i + \sum_{i=0}^{\infty} f(b_i)x^i = f(w(x)) + f(v(x)). \end{aligned}$$

Likewise, we have

$$\begin{aligned} f(w(x)\cdot v(x)) &= f\left(\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(a_i\cdot b_j)x^{i+j}\right) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} f(a_i\cdot b_j)x^{i+j} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} f(a_i)\cdot f(b_j)x^{i+j} \\ &= f(w(x))\cdot f(v(x)). \end{aligned}$$

Thus, f extends to a homomorphism mapping $K[x]$ to $M[x]$. But the kernel of f is just the identity element, since f preserves the degree of any nonzero polynomial. Thus, f extends to an isomorphism from $K[x]$ to $M[x]$, and $f(x) = x$. □

We can apply lemma 13.3 to the case where f is an automorphism on $K[x]$, such as the Frobenius automorphism. By extending the Frobenius automorphism to a polynomial, we can generate irreducible polynomials in $Z_p[x]$. These irreducible polynomials are important, since we can define the field in terms of these polynomials.

PROPOSITION 13.5

Let F be a finite field of characteristic p. For any y in F, let n be the smallest number such that $y^{p^n} = y$. Then

$$g(x) = (x-y)\cdot(x-f(y))\cdot(x-f(f(y)))\cdots(x-f^{n-1}(y))$$

is an irreducible polynomial of degree n in $Z_p[x]$.

PROOF Consider the extension of the Frobenius automorphism onto $F[x]$, as given in lemma 13.3. If we apply this mapping to the polynomial $g(x)$, we get

$$f(g(x)) = (x-f(y))\cdot(x-f(f(y)))\cdot(x-f(f(f(y))))\cdot\cdots\cdot(x-f^n(y)).$$

Recall we picked n to be the smallest number such that $f^n = y$. Thus,

$$f(g(x)) = (x-f(y))\cdot(x-f(f(y)))\cdot(x-f(f(f(y))))\cdot\cdots\cdot(x-f^{n-1}(y))\cdot(x-y),$$

which after rearranging the factors gives us $g(x)$ again.

Since $g(x)$ is fixed by the Frobenius automorphism, each coefficient of $g(x)$ must be fixed by $f(x)$. But the only elements fixed by $f(x)$ are those in Z_p. Thus, $g(x)$ must have all of its coefficients in Z_p, and so is a polynomial in $Z_p[x]$. To show that $g(x)$ is irreducible, suppose that

$$g(x) = h(x)\cdot j(x),$$

where both $h(x)$ and $j(x)$ are polynomials in $Z_p[x]$ of positive degree. Then $f(h(x)) = h(x)$ and $f(j(x)) = j(x)$ since the Frobenius automorphism fixes x and the elements in Z_p. By the unique factorization in $F[x]$, $(x - y)$ has to be a factor of $h(x)$ or $j(x)$, but not both, since $(x - y)$ is a factor of $g(x)$ but $(x-y)^2$ is not. Let us suppose that $h(x)$ has $(x-y)$ as a factor. Any factor of $j(x)$ would have to be a factor of $g(x)$, so such a factor would have the form

$$(x - f^m(y))$$

for some $m > 0$. Thus, $f^m(y)$ is a root of $j(x)$, but y is not. But this is impossible, since $f^m(j(x)) = j(x)$, and so $f^m(j(y)) = j(f^m(y)) = 0$. Therefore, $g(x)$ is an irreducible polynomial in $Z_p[x]$. □

DEFINITION 13.2 The polynomial produced by proposition 13.5 is called the *irreducible polynomial of y over Z_p*. If y is in Z_p, this polynomial is simply $x - y$.

We can now use proposition 13.5 to show us that every finite field can be produced as an extension of Z_p over an irreducible polynomial. While we are at it, we will prove a statement that is true for all fields, not just finite fields.

PROPOSITION 13.6
Let K be any field, and F be a subfield of K. Suppose there is an element y of K such that there are no proper subfields of K containing both F and y. Suppose that there is a polynomial $f(x)$ in $K[x]$ with coefficients in F such that $f(y) = 0$. Suppose further that $f(x)$ is an irreducible polynomial when treated as a polynomial in $F[x]$. Then K is isomorphic to $F[x]/(f(x))$.

PROOF Consider the evaluation homomorphism

$$\phi_y : K[x] \to K$$

restricted on the ring $F[x]$. In other words, we can consider the homomorphism ϕ'_y as the restriction of ϕ_y on $F[x]$. Let us consider the kernel of this homomorphism. Because $f(y) = 0$, $f(x)$ is certainly in the kernel of ϕ'_y. But the kernel cannot be all of $F[x]$, since the constant polynomials are not in the kernel. We know that the kernel is an ideal, and by corollary 12.7, $F[x]$ is a PID, so the kernel can be written as $(g(x))$ for some $g(x)$ in $F[x]$. Yet $f(x)$ is in the kernel, so $g(x)$ divides $f(x)$. But $f(x)$ is irreducible in $F[x]$, and $g(x)$ cannot be a unit, since we have already observed that $(g(x))$ is not all of $F[x]$. Therefore, the kernel of ϕ'_y is $(f(x))$.

From the first ring isomorphism theorem (10.2), the image of ϕ'_y is isomorphic to

$$F[x]/(f(x)).$$

We have already mentioned that $F[x]$ is a PID, so by lemma 12.6 the image is a field. But the field must contain F, since this is the image of the constant polynomials, and also must contain y, the image of the polynomial x. The only subfield of K that contains both y and F is K itself, so $F[x]/(f(x))$ is isomorphic to K. □

One immediate application of proposition 13.6 is to show us that every finite field can be produced as an extension of Z_p over an irreducible polynomial. We will use the polynomial derived in proposition 13.5.

COROLLARY 13.2

For every finite field K of characteristic p, there is an irreducible polynomial $f(x)$ of $Z_p[x]$ such that K is isomorphic to $Z_p[x]/(f(x))$.

PROOF If K is a finite field, by proposition 13.4, the multiplicative group of K^* is cyclic. Thus, there must be an element y that generates K^* as a group. Since K must have finite characteristic p, we will let F be the subfield Z_p. Let $f(x)$ be the irreducible polynomial of y over Z_p given by proposition 13.5.

Even though $f(x)$ is irreducible in $Z_p[x]$, $f(x)$ has $(x - y)$ as a factor when viewed as a polynomial in $K[x]$. Note that since y generates all of K, we see that the conditions for proposition 13.6 are satisfied. Therefore K is isomorphic to $Z_p[x]/(f(x))$. □

We have already seen one field of order 9, produced by the polynomial x^2+1. But there are two other irreducible second degree polynomials in $Z_3[x]$, $x^2 + x + 2$ and $x^2 + 2x + 2$. What if we formed fields using these polynomials? Note that both of these polynomials factor in the field $Z_3[x]/(x^2 + 1)$:

InitDomain[3]
Define[i^2, −1]
K = Ring[{ i }]
Factor[x^2 + x + 2, K]
Factor[x^2 + 2 x + 2, K]

```
gap> InitRing("e","i");
gap> DefineRing("K",[3,3],[[e,i],[i,-e]]);
gap> x := Indeterminate(K,"x");
x
gap> Factor(x^2+x+2,K);
[ x+(2*e+1), x+(2*e+2*i) ]
gap> Factor(x^2 + 2*x + 2, K);
[ x+(e+i), x+(e+2*i) ]
```

Proposition 13.6 hints at what must be happening. The field $Z_3[x]/(x^2 + 1)$ is the smallest field of characteristic 3 for which $x^2 + 1$ factors. But this field

also happens to be the smallest field of characteristic 3 for which $x^2 + x + 2$ and $x^2 + 2x + 2$ factor. This suggests that $Z_3[x]/(x^2 + 1)$, $Z_3[x]/(x^2 + x + 2)$, and $Z_3[x]/(x^2 + 2x + 2)$ are in fact the same field. Could this be so?

The first step in proving this is to find a large field containing both fields.

LEMMA 13.4

Let F and K be two finite fields with the same characteristic p. Then there is a field that contains isomorphic copies of both F and K.

PROOF Since F is a finite field, by corollary 13.2 there is a polynomial $f(x)$ in $Z_p[x]$ such that F is isomorphic to $Z_p[x]/(f(x))$.

Since F and K have the same characteristic, we can consider $f(x)$ to be a polynomial in $K[x]$ as well. Let $g(x)$ be an irreducible factor of $f(x)$ over the domain $K[x]$. Of course, $f(x)$ may already be irreducible in $K[x]$, in which case we let $g(x) = f(x)$.

Now consider the ring $E = K[x]/(g(x))$. Since $K[x]$ is a PID, by lemma 12.6 E is a field. In fact, E contains an element that is a root of the polynomial $g(x)$, namely

$$y = x + (g(x)),$$

since

$$g(y) = g(x + (g(x))\,) = g(x) + (g(x)) = 0 + (g(x)).$$

We can now consider the evaluation homomorphism

$$\phi_y : E[x] \to E.$$

Let us first consider the restriction of this homomorphism to the ring $Z_p[x]$, which we will call ψ. Thus ψ is the homomorphism

$$\psi : Z_p[x] \to E : f(w(x)) = w(y).$$

Since y is a root of $g(x)$ in the field E, and $g(x)$ in turn is a factor of $f(x)$, we see that y is a root of $f(x)$ in the field E. Thus, $f(x)$ is in the kernel of the homomorphism ψ. Since $Z_p[x]$ is a PID, the kernel can be written as $(h(x))$ for some polynomial $h(x)$ in $Z_p[x]$. But since $f(x)$ is in the kernel, $h(x)$ must divide $f(x)$. But $f(x)$ is irreducible, and $h(x)$ cannot be a unit, or else the kernel would be all of $Z_p[x]$, which is impossible since the constant polynomials are not in the kernel. Therefore, the kernel must be $(f(x))$, and so by the first ring isomorphism theorem (10.2), the image of ψ is isomorphic to

$$Z_p[x]/(f(x)),$$

which is in turn isomorphic to F. Thus, there is a subfield of E isomorphic to F.

All we have to do is show that there is a copy of the field K inside of

$$E = K[x]/(g(x)).$$

But we can consider the natural homomorphism

$$i : K[x] \to E$$

given by

$$i(p(x)) = p(x) + (g(x)).$$

If we restrict this homomorphism onto the constant polynomials, we get

$$i' : K \to E.$$

Since $g(x)$ is not a unit, it is clear that the kernel of this homomorphism is just 0. Thus, there is a subfield of E isomorphic to K. Therefore, we have constructed a field that contains isomorphic copies of both F and K as subfields. □

We can now use this lemma to show that there is only one non-isomorphic field of a given order.

COROLLARY 13.3
Any two finite fields of the same order are isomorphic to each other.

PROOF If two fields F and K have the same order, by proposition 13.3, both must have order p^n for some prime number p, and some positive integer n. Thus, both F and K have characteristic p, so by lemma 13.4 there exists a field E that contains isomorphic copies of both F and K as subfields. Let F' and K' be the subfields of E isomorphic to F and K, respectively. Consider the polynomial

$$f(x) = x^{p^n} - x$$

in $E[x]$. Since F' is a subfield of E, the Frobenius automorphism is of order n on this subfield. Thus, every element of F' is a root of $f(x)$. Likewise, every element of K' is also a root of $f(x)$. But by corollary 12.2, $f(x)$ can have at most p^n roots. Thus, the subfields F' and K' must coincide. Hence F' and K' are isomorphic, since they are identical, so F and K must be isomorphic. □

This proposition explains the strange behavior of fields that we discovered in our experiment. Whenever a finite field F is extended though an irreducible polynomial, all irreducible polynomials in $F[x]$ of the same degree factor completely in the new field. The reason is now clear: The field

$$F[x]/(f(x))$$

only depends on the degree of the irreducible polynomial $f(x)$.

We have already seen fields of order 4, 9, and 27 in this chapter. We in fact can refer to them as *the* fields of order 4, 9, or 27. However, there is one question we have yet to answer. Given a prime number p and an integer n, is there a field of order p^n? It seems like all we would need to construct such a field is an irreducible polynomial $f(x)$ in $Z_p[x]$ of degree n, and then the field

$$Z_p[x]/(f(x))$$

would have order p^n. The only problem with this argument is that we have not shown that there *is* an irreducible polynomial of degree n in $Z_p[x]$. In order to construct such irreducible polynomials, we will need to utilize a special class of polynomials—the cyclotomic polynomials. These polynomials have many different uses that crop up in unexpected places.

13.3 Cyclotomic Polynomials

We now pause from our work on finite fields to discuss a special class of polynomials in $\mathbb{Z}[x]$. These polynomials occur in the factorizations of the simple polynomial $x^n - 1$. Although these polynomials are constructed easily, they have a tendency to appear in many different applications, and hence are very useful.

To introduce the cyclotomic polynomials, we will begin by noticing a pattern in the following factorizations:

ClearDefs
Factor[**x**–**1**]
Factor[**x**^**2**–**1**]
Factor[**x**^**3**–**1**]
Factor[**x**^**4**–**1**]
Factor[**x**^**5**–**1**]
Factor[**x**^**6**–**1**]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Factor(x-1,Rationals);
[ x-1 ]
gap> Factor(x^2-1,Rationals);
[ x-1, x+1 ]
gap> Factor(x^3-1,Rationals);
[ x-1, x^2+x+1 ]
gap> Factor(x^4-1,Rationals);
[ x-1, x+1, x^2+1 ]
gap> Factor(x^5-1,Rationals);
[ x-1, x^4+x^3+x^2+x+1 ]
```

```
gap> Factor(x^6-1,Rationals);
[ x-1, x+1, x^2-x+1, x^2+x+1 ]
```

In each factorization there is exactly one polynomial that appears that has not appeared in any previous factorization. Our plan is to find a formula for the irreducible polynomials produced in these factorizations. A natural starting place would be to find all of the complex roots of the polynomial $x^n - 1$. But we have already seen that the primitive n-th roots of unity are of the form e_n^k, where k is coprime to n.

How are the primitive roots of unity related to the factorizations of $x^n - 1$? It is clear that the primitive roots are precisely the complex zeros of $x^n - 1$ that are not zeros of $x^m - 1$ for $m < n$. Thus, if we wish to find the factor of $x^n - 1$ that does not appear in any previous factorizations, we should look for a polynomial whose only complex roots are the primitive n-th roots of unity.

For example, the primitive eighth roots of unity were found to be

$$e_8, \qquad {e_8}^3, \qquad {e_8}^5, \quad \text{and} \quad {e_8}^7.$$

Thus, the simplest polynomial that has these four complex roots would be

InitDomain[0]
e8 = (1/2 + I/2) Sqrt[2]
(x − e8).(x − e8^3).(x − e8^5).(x − e8^7)

```
gap> x := Indeterminate(Rationals,"x");
x
gap> (x-E(8))*(x-E(8)^3)*(x-E(8)^5)*(x-E(8)^7);
x^4+1
```

which simplifies to $x^4 + 1$, which is a factor of $x^8 - 1$. Apparently not only did the imaginary part cancel, but also the square roots simplified. We can use this example for our definition.

DEFINITION 13.3 For $n > 0$, we define the *n-th cyclotomic polynomial* to be the product

$$\Phi_n(x) = (x - {e_n}^{k_1}) \cdot (x - {e_n}^{k_2}) \cdot (x - {e_n}^{k_3}) \cdots (x - {e_n}^{k_i}),$$

where $k_1, k_2, k_3, \ldots, k_i$ are the integers between 0 and n that are coprime to n.

It is sometimes convenient to use a special notation for a product of many terms. Just as the sigma can be used to denote the sum of many terms, a large Π (the upper case π) is used to denote such a product. Thus, we could write

$$\Phi_n(x) = \prod_{\substack{k=1 \\ \text{GCD}(k,n)=1}}^{n} (x - {e_n}^k).$$

In this product, the index k ranges from 1 to n, but we only consider the values of k for which $\mathrm{GCD}(k, n) = 1$. It is apparent from the definition that the degree of the n-th cyclotomic polynomial is $\phi(n)$, where ϕ is Euler's totient function.

Although this definition uses complex numbers, we observed that the polynomials always produced integer coefficients. The next proposition shows us how to find the cyclotomic polynomials without having to work with complex numbers.

PROPOSITION 13.7

For any positive integer n, we have

$$x^n - 1 = \prod_{k|n} \Phi_k(x).$$

Here, the product is taken over all values of k that divide n.

PROOF We will first show that each n-th root of unity is a primitive k-th root of unity for exactly one positive divisor k of n. If $z = {e_n}^s$ is an n-th root of unity, we can let $k = n/\mathrm{GCD}(n, s)$. Then $k \cdot s = n \cdot (s/\mathrm{GCD}(n, s))$ is a multiple of n, so $z^k = 1$. Yet if $z^m = 1$, then $s \cdot m$ must be a multiple of n, so $(s/\mathrm{GCD}(n, s)) \cdot m$ is a multiple of $n/\mathrm{GCD}(n, s)$. But $(s/\mathrm{GCD}(n, s))$ and $(n/\mathrm{GCD}(n, s))$ are coprime, so m would be a multiple of k. Thus, ${e_n}^s$ is a primitive k-th root of unity, with $k = n/\mathrm{GCD}(n, s)$.

Since

$$x^n - 1 = (x - e_n) \cdot (x - {e_n}^2) \cdot (x - {e_n}^3) \cdot \cdots \cdot (x - {e_n}^n),$$

we can collect those factors $(x - {e_n}^s)$ for which ${e_n}^s$ is a primitive k-th root of unity. The result is the formula

$$x^n - 1 = \prod_{k|n} \Phi_k(x).$$

□

To help understand this notation, let us look at the case where $n = 12$. Then proposition 13.7 states that

$$x^{12} - 1 = \prod_{k|12} \Phi_k(x) = \Phi_1(x) \cdot \Phi_2(x) \cdot \Phi_3(x) \cdot \Phi_4(x) \cdot \Phi_6(x) \cdot \Phi_{12}(x).$$

We can observe this factorization using *Mathematica* or GAP.

Factor[x^12 − 1]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Factor(x^12-1, Rationals);
[ x-1, x+1, x^2-x+1, x^2+1, x^2+x+1, x^4-x^2+1 ]
```

Proposition 13.7 at least explains our observation that the factorization of $x^n - 1$ always produces a new factor. However, we have not proven that the cyclotomic polynomials are irreducible in $\mathbb{Z}[x]$. They are indeed irreducible, but we will not need this for our work with finite fields.

COROLLARY 13.4
The n-th cyclotomic polynomial $\Phi_n(x)$ has integer coefficients for all $n > 0$.

PROOF We will prove this using induction on n. Obviously the first cyclotomic polynomial is $x - 1$, which has integer coefficients. Let $n > 1$, and suppose the claim is valid for all previous cyclotomic polynomials. By proposition 13.7, we can find the n-th cyclotomic polynomial as

$$\Phi_n(x) = (x^n - 1)/f(x)$$

where

$$f(x) = \prod_{\substack{k|n \\ k<n}} \Phi_k(x).$$

Since all previous cyclotomic polynomials have integer coefficients, we see by induction that $f(x)$ has integer coefficients. Furthermore, from the definition of the cyclotomic polynomials we see that the leading coefficients must be 1, hence the leading coefficient of $f(x)$ is 1. So by corollary 12.1 the quotient $(x^n - 1)/f(x)$ must in fact have integer coefficients. Therefore, all cyclotomic polynomials have integer coefficients. □

It is actually very easy to generate the n-th cyclotomic polynomial in *Mathematica* or GAP. The commands

Cyclotomic[3, x]
Cyclotomic[6, x]

```
gap> CyclotomicPolynomial(Rationals, 3);
x^2+x+1
gap> CyclotomicPolynomial(Rationals, 6);
x^2-x+1
```

find the third and sixth cyclotomic polynomial, $x^2 + x + 1$ and $x^2 - x + 1$. Notice that the coefficients for these cyclotomic polynomials are either 0 or ± 1. This is the case for $n \leq 100$, but for larger values of n, the coefficients of $\Phi_n(x)$ can be larger. For example, there are two coefficients of -2 in $\Phi_{105}(x)$.

```
gap> CyclotomicPolynomial(Rationals, 105);
x^48+x^47+x^46-x^43-x^42-2*x^41-x^40-x^39+x^36+x^35+x^34+x^33+\
x^32+x^31-x^28-x^26-x^24-x^22-x^20+x^17+x^16+x^15+x^14+x^13+x^\
12-x^9-x^8-2*x^7-x^6-x^5+x^2+x+1
```

The next corollary is another easy consequence of corollary 13.4.

COROLLARY 13.5
If n is divisible by m, with $n > m$, then the polynomial $x^n - 1$ is divisible by $x^m - 1$ in $\mathbb{Z}[x]$. Furthermore, $\Phi_n(x)$ divides

$$\frac{x^n - 1}{x^m - 1}$$

in $\mathbb{Z}[x]$.

PROOF Since n is divisible by m, whenever m is divisible by k, then n is divisible by k. Thus, every factor appearing in

$$x^m - 1 = \prod_{k|m} \Phi_k(x)$$

also appears in

$$x^n - 1 = \prod_{k|n} \Phi_k(x).$$

In fact, the quotient would be the product of the cyclotomic polynomials $\Phi_k(x)$ for which k is a divisor of n, but not of m. Since the cyclotomic polynomials have integer coefficients,

$$\frac{x^n - 1}{x^m - 1}$$

would have integer coefficients. Furthermore, $\Phi_n(x)$ is one of the cyclotomic polynomials in the factorization of $x^n - 1$ which is not in $x^m - 1$. Thus, the n-th cyclotomic polynomial divides $(x^n - 1)/(x^m - 1)$ in $\mathbb{Z}[x]$. □

We now want to find some properties of the cyclotomic polynomials. One of the most important properties is that two different cyclotomic polynomials cannot share a root in the complex numbers. (This is obvious from the definition.) However, we will be working with other fields besides the complex numbers, so we could ask whether a cyclotomic polynomial has multiple roots in *any* field.

DEFINITION 13.4 If r is a root of a polynomial $f(x)$, and $(x - r)^2$ divides $f(x)$, we say r is a *multiple root* of $f(x)$.

We would like to determine when $x^n - 1$ has multiple roots. Our strategy is to discover the form of the quotient

$$\frac{x^n - 1}{x - 1}.$$

For example, $(x^4 - 1)/(x - 1)$ is given by

```
gap> x := Indeterminate(Rationals,"x");
x
gap> (x^4-1)/(x-1);
x^3+x^2+x+1
```

In *Mathematica*, is takes more work to get the answer to simplify.

Expand[Factor[(x^4 – 1)/(x–1)]]

which yields $x^3 + x^2 + x + 1$. By observing other quotients in *Mathematica* or GAP, we can see the general pattern. Using this pattern, we can prove the following lemma.

LEMMA 13.5
If F is any field, then the polynomial $x^n - 1$ has a multiple root if, and only if, n is a multiple of the characteristic of F.

PROOF We first will ask whether 1 is a multiple root of $x^n - 1$. Since 1 is clearly a root,

$$x^n - 1 = (x - 1) \cdot f(x)$$

for some polynomial $f(x)$. But we can use the division algorithm to produce $f(x)$. We claim that

$$f(x) = \sum_{k=0}^{n-1} x^k = 1 + x + x^2 + x^3 + \cdots + x^{n-2} + x^{n-1}.$$

To see this, note that

$$\begin{aligned}(x - 1) \cdot f(x) &= x \cdot f(x) - f(x) \\ &= (x + x^2 + x^3 + \cdots x^{n-1} + x^n) \\ &\quad - (1 + x + x^2 + x^3 + \cdots + x^{n-2} + x^{n-1}) \\ &= x^n - 1.\end{aligned}$$

To see whether 1 is a double root, we observe that

$$f(1) = \sum_{k=0}^{n-1} 1^k = 1 + 1 + 1^2 + 1^3 + \cdots + 1^{n-2} + 1^{n-1} = n.$$

Thus, $f(1)$ is zero if, and only if, n is a multiple of the characteristic of F. Therefore, 1 is a double root of $f(x)$ precisely when the characteristic is positive and divides n.

Now suppose that n is not a multiple of the characteristic, and that r is a double root of $x^n - 1$. Then

$$\frac{x^n - 1}{(x - r)^2}$$

is a polynomial in $F[x]$. If we replace x with $x \cdot r$ we get

$$\frac{(x \cdot r)^n - 1}{(x \cdot r - r)^2} = \frac{x^n r^n - 1}{(x-1)^2 \cdot r^2} = \frac{x^n - 1}{(x-1)^2 \cdot r^2}$$

since $r^n = 1$. However, we have already shown that 1 is not a double root of $x^n - 1$, so the right hand side of this equation cannot be a polynomial. Thus, r is not a double root whenever n is not a multiple of the characteristic. □

This lemma can now be used to generate irreducible polynomials in $Z_p[x]$ of any degree. In fact, these irreducible polynomials are the key to proving that a field of order p^n exists.

PROPOSITION 13.8

Let p be a prime integer, and let $n > 1$. Consider the cyclotomic polynomial

$$\Phi_{(p^n-1)}(x)$$

of order $\phi(p^n - 1)$. Let us consider $g(x)$ to be this polynomial modulo p in $Z_p[x]$. Then $g(x)$ factors in $Z_p[x]$ into irreducible polynomials, all of which have degree n.

PROOF Let $h(x)$ be an irreducible factor of $g(x)$, and let K be the field $Z_p[x]/(h(x))$. We wish to show that the order of K is p^n, since by proposition 13.2 this would indicate that the degree of $h(x)$ is n. Let y be the element

$$y = x + (h(x))$$

in the field K. Then $h(y) = 0$, and hence $g(y) = 0$ in the field K. In fact, $g(x)$ would be a factor of

$$x^{(p^n-1)} - 1,$$

and so $y^{p^n} = y$. In other words, if $f(x)$ is the Frobenius automorphism on K, then $f^n(y) = y$. In fact, $f^n(1) = 1$, and $Z_p[x]$ is generated by x and 1, so we find that $f^n(x) = x$ for all x in K. Thus, the polynomial

$$x^{p^n} - x$$

has at least $|K|$ roots. By corollary 12.2, $|K|$ can have at most p^n elements.

To show that $|K| = p^n$, let us suppose that $|K| = p^m$, where $m < n$. Then m is the smallest number for which $f^m(x) = x$ for all x in K. It is clear that m would have to divide n, since $f^n(x)$ is also x for all x in K.

Since $f^m(y) = y$, we see that y is a root of the polynomial

$$x^{(p^m-1)} - 1.$$

By corollary 13.5, $\Phi_{(p^n-1)}(x)$ divides

$$\frac{x^{(p^n-1)}-1}{x^{(p^m-1)}-1}$$

in $\mathbb{Z}[x]$, since (p^m-1) divides (p^n-1). Thus, in $Z_p[x]$, $g(x)$ divides

$$\frac{x^{(p^n-1)}-1}{x^{(p^m-1)}-1}.$$

Since $g(y)=0$, and also $y^{(p^m-1)}=1$, we see that y would be a multiple root of $x^{(p^n-1)}-1$. But by lemma 13.5, this polynomial can only have a multiple root if (p^n-1) is a multiple of p, which it clearly isn't. Thus, $m=n$, and so $|K|=p^n$. By proposition 13.2, the irreducible factors of $g(x)$ over $Z_p[x]$ all have degree n. ∎

We can now prove what we had suspected was true from the experiments: that there is precisely one field of order p^n, where $n>0$ and p is a prime number.

COROLLARY 13.6
If p is a prime number, and n is a positive integer, there exists a unique field (up to isomorphism) of order p^n.

PROOF We have already shown in corollary 13.3 that finite fields of the same order are isomorphic, so all we have to show is that there is a field of order p^n. By proposition 13.8, the cyclotomic polynomial

$$\Phi_{(p^n-1)}(x)$$

factors in $Z_p[x]$ into irreducible factors of degree n. If we let $A(x)$ be one of those irreducible factors, then by proposition 13.2, the field

$$K=Z_p[x]/(A(x))$$

has order p^n. ∎

DEFINITION 13.5 If $q=p^n$, where p is prime and $n>0$, then the *Galois field of order q,* denoted $GF(q)$, is the unique field of order q given in corollary 13.6.

For example, the *official* name for the "complex numbers modulo 3" we have been working with is $GF(9)$. Whenever p is prime, we can write $GF(p)$ for the field Z_p.

We can enter finite fields into GAP using this notation. For example, the faster way to enter $GF(9)$ in GAP is

```
gap> K := GF(9);
GF(3^2)
gap> List(K);
[ 0*Z(3), Z(3)^0, Z(3), Z(3^2), Z(3^2)^2, Z(3^2)^3, Z(3^2)^5,
  Z(3^2)^6, Z(3^2)^7 ]
```

A bit of explanation is in order here. We have established in proposition 13.4 that the multiplicative group is cyclic, so we can let $Z(9)$ be a generator of the multiplicative group, so that all nonzero elements can be expressed as a power of $Z(9)$. In GAP 3, the elements were listed as

```
[ 0*Z(9), Z(9)^0, Z(9), Z(9)^2, Z(9)^3, Z(9)^4, Z(9)^5, Z(9)^6,
  Z(9)^7, Z(9)^8 ]
```

but this causes a problem in that $GF(3)$ should be automatically a subgroup of $GF(9)$. Hence, $Z(9)^4$ should simplify to $Z(3)$, and the multiplicative identity is listed as $Z(3)^0$ instead of $Z(9)^0$.

If we list the elements in the order of increasing powers of $Z(9)$, the multiplication table becomes easy to understand.

```
gap> L := [0*Z(9), Z(9)^0, Z(9), Z(9)^2, Z(9)^3, Z(9)^4, Z(9)^5,
> Z(9)^6, Z(9)^7 ];
[ 0*Z(3), Z(3)^0, Z(3^2), Z(3^2)^2, Z(3^2)^3, Z(3), Z(3^2)^5,
  z(3^2)^6, Z(3^2)^7 ]
gap> NumberElements := true;
true
gap> MultTable(L);

*        |1  2  3  4  5  6  7  8  9
---------+------------------------
0*Z(3)   |1  1  1  1  1  1  1  1  1
Z(3)^0   |1  2  3  4  5  6  7  8  9
Z(3^2)   |1  3  4  5  6  7  8  9  2
Z(3^2)^2 |1  4  5  6  7  8  9  2  3
Z(3^2)^3 |1  5  6  7  8  9  2  3  4
Z(3)     |1  6  7  8  9  2  3  4  5
Z(3^2)^5 |1  7  8  9  2  3  4  5  6
Z(3^2)^6 |1  8  9  2  3  4  5  6  7
Z(3^2)^7 |1  9  2  3  4  5  6  7  8
```

Except for the zero element, we have diagonal streaks of elements in the multiplication table, indicative of a cyclic group. What is not so self-explanatory is the addition table.

```
gap> AddTable(L);

+        |1  2  3  4  5  6  7  8  9
---------+------------------------
0*Z(3)   |1  2  3  4  5  6  7  8  9
Z(3)^0   |2  6  4  9  8  1  5  7  3
Z(3^2)   |3  4  7  5  2  9  1  6  8
Z(3^2)^2 |4  9  5  8  6  3  2  1  7
Z(3^2)^3 |5  8  2  6  9  7  4  3  1
Z(3)     |6  1  9  3  7  2  8  5  4
Z(3^2)^5 |7  5  1  2  4  8  3  9  6
Z(3^2)^6 |8  7  6  1  3  5  9  4  2
Z(3^2)^7 |9  3  8  7  1  4  6  2  5
```

The addition table is hard to understand because we have yet to determine *which* of the generators GAP assigned to $Z(9)$. In other words, we must determine which irreducible polynomial of degree 2 over Z_3 should be used to define the field. There are in fact six such polynomials: x^2+1, x^2+x+2, x^2+2x+2, $2x^2+2$, $2x^2+2x+1$, and $2x^2+x+1$. This list can be reduced to three polynomials if we insist that the leading coefficient be 1. But if we use x^2+1 for the defining polynomial, as we did for the "complex numbers mod 3," then the roots of this polynomial, $\pm i$, would not be generators of the multiplicative group, and hence could not be used to define $Z(9)$.

DEFINITION 13.6 A polynomial $f(x)$ over a finite field F is a *privative polynomial* if it is irreducible, has a leading coefficient of 1, and $x+(f(x))$ is a multiplicative generator of the finite field $F[x]/(f(x))$.

Although we can rule out using x^2+1 to define $Z(9)$, there are still two primitive polynomials x^2+x+2 and x^2+2x+2. The roots of these two polynomials in $GF(9)$ are all multiplicative generators. Both of these polynomials will work equally well, so we define the following "tie breaker."

DEFINITION 13.7 The *Conway polynomial* of degree n over Z_p is the primitive polynomial of degree n in $Z_p[x]$ for which

1. The polynomial is compatible with the way that the subfields of $GF(p^n)$ are defined. To be compatible, for all divisors d of n less than n, the $(\frac{p^n-1}{p^d-1})$th power of the zeros of the polynomial must be zeros of the Conway polynomial of degree d over Z_p.
2. If two or more primitive polynomials satisfy the compatibility condition, let d be the highest power of x for which the coefficients differ. If $n-d$ is even, pick the one with the smallest coefficient from the set $\{0,1,\ldots p-1\}$. If $n-d$ is odd, pick the largest, unless there is one with a coefficient of 0.

This definition at first seems counter-intuitive. Logically, a zero coefficient is always preferred over a nonzero term, but sometimes we pick the polynomial with the largest coefficient, and sometimes use the one with the smallest. But to understand why this is so, consider the first degree Conway polynomials. Since all of the primitive polynomials are of the form $x+c$, with $c \neq 0$, they differ only in the constant term. Hence $d=0$, so $n-d$ will be odd, and we should select the primitive polynomial with the largest c. This in turn will make the root of this polynomial be as *small* as possible. So for p prime, `Z(p)` will represent the smallest generator of the group Z_p^*. For example, `Z(5)` will be GAP's way of representing 2 in the field Z_5, and `Z(7)` will represent 3 in the field Z_7. In general, the Conway polynomial is designed so that the roots will be minimized.

Let us use this definition to find the Conway polynomial of degree 2 over Z_3. In order to understand the compatibility condition, we must first find the Conway polynomial of degree 1 over Z_3. Since there is only one generator of Z_3, namely 2, there is only one primitive polynomial of degree 1, $x-2 = x+1$.

Now in order for a primitive polynomial of degree 2 to be compatible, the 4th power of the roots must be a root of $x+1$ ($(3^2-1)/(3^1-1) = 4$). But the 4th power of all four generators in $GF(9)$ produces 2, so both x^2+x+2 and x^2+2x+2 satisfy the compatibility condition, but x^2+1 does not, since $i^4 = 1 \neq 2$ in $GF(9)$.

Of the two possible primitive polynomials remaining, we look for the largest power of x for which these differ, (x^1), and since $n-d = 1$ is odd, and neither x^1 coefficient is 0, we pick the larger of the two possible coefficients. So the Conway polynomial is x^2+2x+2.

GAP has many Conway polynomials precomputed, since they are time consuming to compute from scratch. These Conway polynomials $f(x)$ are then used to define $GF(p^n) = Z_p[x]/(f(x))$.

```
gap> x := Indeterminate(GF(3),"x");
x
gap> ConwayPolynomial(3,2);
x^2-x-Z(3)^0
```

GAP expresses the polynomial in terms of $Z(p)$, so this is $x^2 - x - 1 = x^2 + 2x + 2$. Thus, if we define the field $Z_3[x]/(x^2+2x+2)$, and order the elements in powers of the generator,

```
gap> InitRing("e","a");
gap> DefineRing("K",[3,3],[[e,a],[a,a+e]]);
gap> L := [0*a, e, a, a^2, a^3, a^4, a^5, a^6, a^7 ];
[ 0*e, e, a, e+a, e+2*a, 2*e, 2*a, 2*e+2*a, 2*e+a ]
gap> NumberElements := true;
true
gap> AddTable(L);

+       |1  2  3  4  5  6  7  8  9
--------+------------------------
0*e     |1  2  3  4  5  6  7  8  9
e       |2  6  4  9  8  1  5  7  3
a       |3  4  7  5  2  9  1  6  8
e+a     |4  9  5  8  6  3  2  1  7
e+2*a   |5  8  2  6  9  7  4  3  1
2*e     |6  1  9  3  7  2  8  5  4
2*a     |7  5  1  2  4  8  3  9  6
2*e+2*a|8  7  6  1  3  5  9  4  2
2*e+a   |9  3  8  7  1  4  6  2  5
```

we find that the pattern of the addition table matches that of the addition table for $GF(9)$. Of course the multiplication tables would also have the same pattern, since both are defined in terms of a generator.

Mathematica also has the ability to find Conway polynomials, but the routine is much slower than GAP's, since they are not precomputed.

ConwayPolynomial[3, 2, x]
$2 + 2x + x^2$

The Galois fields have many applications. A code very similar to the RSA code studied in chapter 3 of group theory was developed using Galois fields of characteristic 2. For a long time the field of order 2^{127} was used, since the multiplicative group is of order $2^{127} - 1$, which happens to be prime. (Primes of this form are called Mersenne primes.) This code had the advantage that the key was much shorter than the RSA key, and multiplication in this field could be quickly implemented in binary hardware. However, due to the special properties of finite fields, this code was recently cracked. In order to ensure safety of the encryption, the size of the field had to be upped to order 2^{2201}, which diminished the advantage over the RSA code.

But there is another type of code based on Galois fields, called the Reed-Solomon code, which is not used for security but rather for the storage or transfer of digital data. All digital information, such as the storage of a file in a computer or a song on a compact disc, is stored as a string of "bits" that are either 0 or 1. We will let K denote a finite field of characteristic 2. For example, if $K = GF(256)$, then each element of K would correspond to a computer "byte." (Each byte is eight bits.) A string of n bytes $(a_0, a_1, a_2, a_3, \ldots, a_{n-1})$ is encoded as a polynomial in K:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots a_{n-1} x^{n-1}.$$

The encryption of this list of elements is simply the evaluation of this polynomial at the 256 elements of K. That is, if g is a generator of the multiplicative group K^*, then

$$f(0),\ f(g),\ f(g^2),\ f(g^3),\ \ldots,\ f(g^{255})$$

is transmitted in place of the numbers $a_0, a_1, a_2, \ldots a_{n-1}$. We know from corollary 12.3 that we can reconstruct the original list of elements from any n of the numbers transmitted. Thus, if there are some errors in the transmission, the original list can still be determined. Using combinatorial reasoning, Reed and Solomon showed that as many as $(255 - n)/2$ errors could occur, and yet the original list of elements can be decoded.

For example, if $n = 251$, then every 251 bytes is converted to a 250 degree polynomial, which is evaluated at the 256 elements of K. Even if two of these bytes are transmitted incorrectly, the 251 original bytes can be correctly reconstructed. This is an example of what is called an "error-correcting code." This code was used by the *Voyager II* spacecraft to transmit pictures of Uranus and Neptune back to Earth. [16] A version of this code (using a larger field K) is used to store the digital music on a compact disc. Current CD players can cope with errors as long as 4000 consecutive bits on the CD, typically caused by a scratch on the CD surface. The Reed-Solomon code also allows over 500 channels of digital television.

The ironic part of this code is that, when Reed and Solomon first discovered the code in 1960, [15] it was described as "interesting, but probably not

practical." It wasn't until hardware technology advanced to the point that the code could be implemented before the real value of this code was evident. As with most mathematics, the usefulness of a particular result is not seen until long after the result is published.

One final application of finite fields arises from the study of simple groups. Almost all of the simple groups besides the alternating groups are the Chevalley groups, which are defined in terms of finite fields. For example, the simple group $\mathrm{Aut}(Z_{24}^*)$ can be expressed as the 3 by 3 matrices in the field Z_2 with determinant 1. This example can be generalized to a group G of m by m matrices over any finite field of order p^n. When $p^n > 2$, there may be a nontrivial center Z formed by diagonal matrices. However, we can form the quotient group G/Z. The group generated, denoted $L_m(p^n)$, will be simple if $m > 2$, or if $m = 2$ and $p^n > 3$. [9, p. 223]

There are several other ways of forming simple groups using finite fields. In fact, besides the alternating groups, there are only 26 finite simple groups that are not expressed using finite fields. Thus, finite fields are of key importance in the classification of all finite simple groups.

13.4 Finite Skew Fields

Since we have completely classified all finite fields, a natural question is whether we can classify all finite skew fields, and whether these can be easily entered into *Mathematica*. At first this seems like it would be a harder problem, since there are many non-abelian groups, and many non-commutative rings. However, a surprising result is that there are *no* finite skew fields. In this section we will prove this remarkable result, known as Wedderburn's theorem.

We begin by carrying over some ideas from group theory. One of the ways we studied non-abelian groups was to find the center of the group, since this was always a normal subgroup. We can ask whether the set of elements of a skew field that commute with all of the elements forms a special set.

DEFINITION 13.8 Let K be a skew field. Then the set of all elements x of K such that $x \cdot y = y \cdot x$ for all $y \in K$ is called the *center* of K.

Let us look at an example. The only skew field we have seen is the ring of quaternions, $\mathbb{H}$. The *Mathematica* command

InitQuaternions

allows us to experiment with this skew field. What is the center of this skew field? To answer this question, let us first define two typical elements in $\mathbb{H}$.

A = u0 + u1 I + u2 J + u3 K
B = v0 + v1 I + v2 J + v3 K

These will commute as long as $A \cdot B - B \cdot A = 0$. By computing

A.B - B.A

to be

$$-2u_2v_1k + 2u_3v_1j + 2u_1v_2k - 2u_3v_2i - 2u_1v_3j + 2u_2v_3i,$$

the only way that this could be zero for all v_1, v_2, and v_3 is for $u_1 = u_2 = u_3 = 0$. Thus, the center of $\mathbb{H}$ is basically the field of real numbers. (Since GAP only works with indeterminates over a commutative ring, this computation can only be done in *Mathematica*.)

LEMMA 13.6
The center of a skew field forms a field.

PROOF Let K be a skew field, and let Z be its center. We first will show that Z is a subring. If x and y are two elements in Z, and k is any element in K, then

$$(x - y) \cdot k = x \cdot k - y \cdot k = k \cdot x - k \cdot y = k \cdot (x - y)$$

and

$$(x \cdot y) \cdot k = x \cdot (y \cdot k) = x \cdot (k \cdot y) = (x \cdot k) \cdot y = (k \cdot x) \cdot y = k \cdot (x \cdot y).$$

Thus, both $x - y$ and $x \cdot y$ are in Z. By proposition 10.1, Z is a subring of K.

Both 0 and the identity element are obviously in Z, so Z is nontrivial. Since Z is commutative, all we have left to prove is that every nonzero element of Z is invertible. If $x \neq 0$ is an element in Z and k is in K, then $x \cdot k = k \cdot x$. The inverse of x exists in K, so we can multiply both sides of the equation on both the left and the right by x^{-1}:

$$x^{-1} \cdot (x \cdot k) \cdot x^{-1} = x^{-1} \cdot (k \cdot x) \cdot x^{-1}.$$

Thus,

$$k \cdot x^{-1} = x^{-1} \cdot k$$

for all k in K, and so x^{-1} is in the center Z. Thus, Z is a field. □

Another concept from group theory that carries over into the study of fields is the normalizer. Recall the definition of a normalizer of a subset S of a group G. We defined

$$N_G(S) = \{g \in G \quad | \quad g \cdot S \cdot g^{-1} = S\}.$$

We would like to apply the normalizer to the multiplicative group of a field. In particular, we would like to consider the normalizer of a particular element, that is, when $S = \{y\}$.

Let us find the normalizer of the element I in the nonzero quaternions. This consists of all elements A such that $A \cdot I \cdot A^{-1} = I$. The *Mathematica* command

Simplify[A.I.(A^(−1)) − I]

shows that these are equal whenever

$$\frac{2((u_1u_2 + u_0u_3)j + (-u_0u_2 + u_1u_3)k - i(u_2^2 + u_3^3))}{u_0^2 + u_1^2 + u_2^2 + u_3^2}$$

is zero, which can only happen if $u_2 = u_3 = 0$. In fact, if A is nonzero, this is sufficient, so we see that the normalizer of i is the set of nonzero elements of the form $u_0 + u_1 i$.

The normalizer does not quite form a field, since it does not include the zero element. Yet if we added the zero element to $N_{\mathbb{H}}(I)$, we get a field equivalent to the complex numbers. It is not hard to show that for any skew field, whenever we add the zero element to the normalizer, we will either get a field or a skew field.

LEMMA 13.7

Let K be a skew field, and let k be an element of K. Then if we let

$$Y_k = \{0\} \cup N_{K^*}(k),$$

then Y_k is a division ring containing the center of K.

PROOF Let us begin by rewriting the set Y_k. Because

$$N_{K^*}(\{k\}) = \{x \in K^* \mid x \cdot k \cdot x^{-1} = k\},$$

we can simply say $N_{K^*}(\{k\})$ consists of all elements of K^* such that $x \cdot k = k \cdot x$. Of course 0 satisfies this equation as well, so we can write

$$Y_k = \{x \in K \mid x \cdot k = k \cdot x\}.$$

When written in this form, it is obvious that the center is in Y_k. Furthermore, if x and y are in Y_k, then

$$(x - y) \cdot k = x \cdot k - y \cdot k = k \cdot x - k \cdot y = k \cdot (x - y)$$

and

$$(x \cdot y) \cdot k = x \cdot (y \cdot k) = x \cdot (k \cdot y) = (x \cdot k) \cdot y = (k \cdot x) \cdot y = k \cdot (x \cdot y).$$

Thus, by proposition 10.1, Y_k is a subring of K.

Finally, if x is a nonzero element in Y_k, then $x \cdot k = k \cdot x$. Thus,

$$x^{-1} \cdot (x \cdot k) \cdot x^{-1} = x^{-1} \cdot (k \cdot x) \cdot x^{-1},$$

so

$$k \cdot x^{-1} = x^{-1} \cdot k.$$

Thus, every nonzero element of Y_k is invertible, so Y_k is a division ring. □

We now can apply the center and normalizer to *finite* division rings. We first need a lemma that will help us out regarding the divisibility of the orders of finite fields.

LEMMA 13.8

Let y, n, and m be positive integers, with $y > 1$. Then

$$\frac{y^n - 1}{y^m - 1}$$

is an integer if, and only if, n is divisible by m. Furthermore, if n is divisible by m, with $n > m$, then

$$\frac{y^n - 1}{y^m - 1}$$

is divisible by the number $\Phi_n(y)$.

PROOF First suppose that n is divisible by m. Then by corollary 13.5, $x^m - 1$ divides $x^n - 1$, and in fact $\Phi_n(x)$ divides

$$\frac{x^n - 1}{x^m - 1}.$$

Note that since $y > 1$, $y^m > 1$, so $y^m - 1 > 0$. Thus, y is not a root of $x^m - 1$, so we can apply the evaluation homomorphism ϕ_y and find that

$$\frac{y^n - 1}{y^m - 1}$$

is divisible by $\Phi_n(y)$.

Now suppose that n is not divisible by m. Then $n = m \cdot k + p$ for some $0 < p < m$. But note that

$$y^n - 1 = y^{(m \cdot k + p)} - 1 = y^{m \cdot k} \cdot y^p - 1 = y^p(y^{m \cdot k} - 1) + y^p - 1.$$

Thus,

$$\frac{y^n - 1}{y^m - 1} = y^p \cdot \frac{y^{m \cdot k} - 1}{y^m - 1} + \frac{y^p - 1}{y^m - 1}.$$

We have already seen that $y^{(m \cdot k} - 1)/(y^m - 1)$ is an integer, but $y^p < y^m$, so the last term cannot possibly be an integer. Therefore, $(y^n - 1)/(y^m - 1)$ is an integer if, and only if, n is a multiple of m. □

This lemma reveals the possible orders of division rings within a finite division ring.

COROLLARY 13.7

Let K be a finite division ring of order p^n, and let F be a subring that is a division ring of order p^m. Then n is a multiple of m.

PROOF Consider the multiplicative groups K^* and F^*. Certainly F^* is a subgroup of K^*, since F is a subring of K. Notice that K^* contains $p^n - 1$ elements, while $|F^*| = p^m - 1$. By Lagrange's theorem (3.1), $p^m - 1$ must be a factor of $p^n - 1$. So by lemma 13.8, n must be a multiple of m. □

Note that this corollary has applications in finite fields. For example, it shows that the field of order 16 cannot have a subfield of order 8.

There is one more tool that we need from group theory, which stems from the normalizer. We discovered in section 7.4 that the class equation was a powerful tool in analyzing groups. In fact, all three Sylow theorems hinge on the class equation. So let us observe how this tool applies to skew fields. Recall that the class equation theorem (7.2) stated that when G is a finite group, then

$$|G| = \sum_g \frac{|G|}{|N_G(\{g\})|}$$

where the sum runs over one g from each conjugacy class.

If K is a finite skew field, we can apply the class equation theorem to the multiplicative group K^*, and find that

$$|K^*| = \sum_k \frac{|K^*|}{|N_{K^*}(\{k\})|}.$$

We can make the obvious substitutions $|K^*| = |K| - 1$, and $|N_{K^*}(\{k\})| = |Y_k| - 1$. The equation now looks like

$$|K| - 1 = \sum_k \frac{|K| - 1}{|Y_k| - 1}$$

where the sum runs from one k from each conjugacy class of K^*.

We are almost ready to use the class equation to prove that finite skew field cannot exist. But first we need to prove a simple inequality about the evaluation of a cyclotomic polynomial at a positive integer.

LEMMA 13.9
If $n > 1$, then the cyclotomic polynomial evaluated at $y \geq 2$, $\Phi_n(y)$, is greater than $y - 1$.

PROOF From the definition,

$$\Phi_n(x) = \prod_{\substack{k=1 \\ \text{GCD}(k,n)=1}}^{n} (x - {e_n}^k).$$

Plugging in $x = y$, and taking the absolute value of both sides, we get

$$\begin{aligned} |\Phi_n(y)| &= \prod_{\substack{k=1 \\ \text{GCD}(k,n)=1}}^{n} |y - {e_n}^k| \\ &> \prod_{\substack{k=1 \\ \text{GCD}(k,n)=1}}^{n} (y-1) \geq (y-1). \end{aligned}$$

Here, the inequality $|y - (e_n)^k| > (y - 1)$ comes from the fact that real part of ${e_n}^k$ is less than 1 when $n > 1$. □

The final step is to use lemma 13.9 to prove a contradiction in the class equation for finite skew fields.

THEOREM 13.2: Wedderburn's Theorem
There are no finite skew fields.

PROOF Suppose that K is a finite skew field. By proposition 13.3 K is of order p^m for some prime p and some $m > 0$. Let Z be the center of K. Since Z is a subring of K which is a field, by corollary 13.7, Z is of order $y = p^a$, where $m = n \cdot a$ for some $n > 0$. Thus, $|K| = p^{n \cdot a} = y^n$. Note that since K is a skew field, n must be greater than 1. We have from the class equation theorem (7.2)

$$|K| - 1 = \sum_k \frac{|K| - 1}{|Y_k| - 1},$$

where the sum runs from one k from each conjugacy class of K^*. Note that when k is in Z^*, k is in its own conjugacy class, and $Y_k = K$. Thus, the terms in the sum corresponding to elements in Z^* are equal to 1. There are of course $|Z^*| = y - 1$ such terms. For the other terms in the sum, Y_k is a proper subring of K that contains Z. By lemma 13.7, Y_k is a division ring, and so by corollary 13.7, $|Y_k| = y^r$ for some r which is a factor of n. If we let

$w = \Phi_n(y)$ we see by lemma 13.8 that w divides the term

$$\frac{|K|-1}{|Y_k|-1} = \frac{y^n-1}{y^r-1}.$$

Furthermore, w divides the left hand side of the class equation, $|K|-1$. In fact, the only terms in the class equation that are not divisible by w are the $y-1$ terms that are equal to 1, coming from the invertible elements of the center Z. Thus, $y-1$ must be divisible by w. But this is impossible, since $y-1 < w$ by lemma 13.9, for $n > 1$. This contradiction proves that finite skew fields cannot exist. □

In a sense, the non-existence of finite skew fields is sad, since there would have been plenty of applications for finite skew fields in cryptography and group theory had they existed. On the other hand, this result, when combined with the classification of all finite fields, means that we have found all finite division rings.

Problems for Chapter 13

Interactive Problems

13.1 The polynomial $x^4 + x + 1$ is irreducible in the field Z_2. Use this polynomial to define a field of order 16 in *Mathematica* or GAP. Show that there is a subfield of order 4 in this field. Is there a subfield of order 8 in this field?

13.2 First define the field Z_2 in *Mathematica* or GAP,

InitDomain[2]
Z2 = {0, 1}

```
gap> InitRing("e");
gap> DefineRing("Z2",[2],[[e]]);
gap> x := Indeterminate(Z2,"x");
x
```

and then show that the cyclotomic polynomial $\Phi_{(2^3-1)}(x)$ factors in the field Z_2 into irreducible polynomials of degree 3. Show by process of elimination that the only irreducible polynomials of degree 3 are the ones given in this factorization.

13.3 First define the field Z_2 in GAP or *Mathematica* as in problem 13.2. Then show that the cyclotomic polynomial $\Phi_{(2^4-1)}(x)$ factors in the field Z_2

into irreducible polynomials of degree 4. Find one more irreducible polynomial of degree 4 besides the ones given in this factorization.

Hint: Factor the polynomial $x^{2^4} - x$.

13.4 First define the field Z_2 in GAP or *Mathematica* as in problem 13.2. Then show that the cyclotomic polynomial $\Phi_{2^5-1}(x)$ factors in the field Z_2 into irreducible polynomials of degree 5. Does this factorization give all of the irreducible polynomials of degree 5 over Z_2?

13.5 First define the field Z_3 in *Mathematica* or GAP:

InitDomain[3]
Z3 = {0, 1, 2}

```
gap> InitRing("e");
gap> DefineRing("Z3",[3],[[e]]);
gap> x := Indeterminate(Z3,"x");
x
```

and then show that the cyclotomic polynomial $\Phi_{3^2-1}(x)$ factors in the field Z_3 into irreducible polynomials of degree 2. What irreducible quadratic polynomial in Z_3 have we seen that is not in the list of factors?

13.6 First define the field Z_3 as in problem 13.5. Then find the factorization of the polynomial $x^{3^3} - x$. Show that all irreducible polynomials with leading term of x^3 are in this factorization. For an explanation see problem 13.26.

13.7 *Mathematica* can be used to explore skew fields besides $\mathbb{H}$. Consider the following ring of characteristic 0:

InitRing
Define[a^3, 3 a+1]
Define[b^3, 2]
Define[b.a, 2 b – a.a.b]

This produces a ring that is a 9-dimensional extension of $\mathbb{Q}$. A basis for this ring would be $\{1, a, a^2, b, a \cdot b, a^2 \cdot b, b^2, a \cdot b^2, a^2 \cdot b^2\}$. If

w1 = C[1] + C[2] a + C[3] a.a
w2 = C[4] + C[5] a + C[6] a.a
w3 = C[7] + C[8] a + C[9] a.a
w = w1 + w2.b + w3.b.b

then w is the general element of this ring. To show that this ring is in fact a skew field for rational values of $C_1, C_2, \ldots C_9$, perform the following operations:

v1 = b.w1.b.w1.b − 2 b.w2.b.w3.b
v2 = 2 w3.b.b.w3.b − w2.b.b.w1.b
v3 = w2.b.w2.b.b − w3.b.w1.b.b
v = Expand[v1 + v2.b + v3.b.b]
R = v.w

Using this value of R, find a formula for w^{-1}. Can you prove that R is never zero if $C_1, C_2, C_3, \ldots C_9$ are rational?

Hint: If $R = 0$ for rational values of $C_1, \ldots C_9$, we can multiply by the common denominator to find a solution to $R = 0$ for integer values. In fact, we may assume that $C_1, C_2, C_3, \ldots C_9$ have no common factors. Show that the first three constants must be even. After a substitution, show that C_4, C_5, C_6 must be even. After yet another substitution, show that the remaining constants are even, leading to a contradiction.

13.8 Use *Mathematica* or GAP to find the Conway polynomial of degree 6 over Z_2. Show that raising a root of this polynomial to the 9th power produces a zero of the Conway polynomial of degree 3 over Z_2, and raising this root to the 21st power produces a zero of the Conway polynomial of degree 2 over Z_2. Hence, the compatibility condition is satisfied.

Non-Interactive Problems

13.9 The polynomial $x^2 + x + 1$ is irreducible in the field Z_2. Write out by hand the addition and multiplication tables of the field $Z_2[x]/(x^2 + x + 1)$.

Hint: There are only four elements.

13.10 The polynomial $x^3 + x + 1$ is irreducible in the field Z_2. Write out by hand the addition and multiplication tables of the field $Z_2[x]/(x^3 + x + 1)$.

13.11 The polynomial $x^2 + x + 2$ is irreducible in the field Z_3. Write out by hand the addition and multiplication tables of the field $Z_3[x]/(x^2 + x + 2)$.

13.12 Construct addition and multiplication tables for a field with 16 elements.

13.13 Using table 11.2 in chapter ch:intdomain of the field of "complex numbers modulo 3," find the generators of the multiplicative group of this field.

13.14 By proposition 13.4, the nonzero elements of Z_p form a cyclic group under multiplication. Any generator of this group is called a *primitive root* of p. Find the primitive roots of the primes 17, 23, and 31. For a given prime, how many primitive roots will there be?

13.15 Show that if F is a field of characteristic p, and x is a generator of the multiplicative group, then x^p is also a generator of the multiplicative group.

13.16 If p is a prime number of the form $4n+1$, show that there is a solution to the equation

$$x^2 \equiv -1 \pmod{p}.$$

Hint: By proposition 13.4, Z_p^* is isomorphic to Z_{p-1}. A solution to the equation would have order 4.

13.17 Use problem 13.16 to show that a prime of the form $4n + 1$ is not prime in the domain $Z[i]$.

Hint: Let x be the solution to the equation in problem 13.16. What is $(x + i)(x - i)$?

13.18 Use problem 13.17 to prove the two square theorem of Fermat: Every prime number of the form $4n + 1$ can be expressed as the sum of two squares.

Hint: Since p is not prime in the domain $Z[i]$, and $Z[i]$ is a UFD, p is reducible in $Z[i]$. If $a + bi$ is one factor, what is the other factor?

13.19 Let F be a field of prime characteristic p. Show that the intersection of all of its subfields of F is a field of order p.

13.20 Let F be a finite field of characteristic p. Show that $F(x)$, the field of quotients of the polynomial ring $F[x]$, is an infinite field of characteristic p.

13.21 Let F be any field. Show that no two finite subfields of F can have the same number of elements.

Hint: See the proof for corollary 13.3.

13.22 Let F be a field of order p^n. Show that if K is a subfield of F then K has order p^d for some number d that divides n.

13.23 Let F be a field of order p^n. Show that if d divides n, then there is a unique subfield of order p^d.

Hint: See problem 13.21 for the uniqueness part.

13.24 Let p be prime and $f(x)$ an irreducible polynomial of degree 2 in $Z_p[x]$. If K is a finite field of order p^3, show that $f(x)$ is also irreducible in $K[x]$.

13.25 Prove that the group of automorphisms of a field of order p^n is isomorphic to Z_n. That is, prove that there are no other automorphisms other than the ones generated by the Frobenius automorphism.

13.26 Let p be a prime number. Show that every irreducible polynomial with a leading term of x^n in the field Z_p is found in the factorization of the polynomial $x^{p^n} - x$.

Hint: If $f(x)$ is an irreducible polynomial of degree n, then $Z_p[x]/(f(x))$ is the Galois field $GF(p^n)$. Show that every element in this field is a root of the polynomial $x^{p^n} - x$. Therefore, the roots of $f(x)$ in the field $GF(p^n)$ are also roots of $x^{p^n} - x$.

For problems **13.27** through **13.30**: Find the cyclotomic polynomial.

13.27 $\Phi_6(x)$ **13.28** $\Phi_9(x)$ **13.29** $\Phi_{10}(x)$ **13.30** $\Phi_{13}(x)$

13.31 Prove that the constant coefficient of the n-th cyclotomic polynomial $\Phi_n(x)$ is equal to -1 when $n = 1$, and is 1 when $n > 1$.

Hint: Use induction along with proposition 13.7.

13.32 Prove that the n-th cyclotomic polynomial $\Phi_n(x)$ is a "palindrome polynomial" when $n > 1$. That is, the list of coefficients read the same going forward or backward.

Hint: Whenever x is a primitive n-th root of unity, x^{-1} will also be a primitive n-th root. What happens if we replace x with $1/y$ in the polynomial? You may use the result of problem 13.31.

13.33 Prove that if p is a prime, and $n > 0$, then

$$\Phi_{p^n}(x) = \Phi_p(x^{p^{n-1}}).$$

13.34 Prove that $\phi(p^n - 1)$ is divisible by n, where ϕ is Euler's totient function.

Hint: See proposition 13.8.

13.35 Prove that the primitive polynomials of degree n over Z_p are precisely the factors of $\Phi_{p^n-1}(x)$ over the field Z_p.

13.36 Prove that every element in a finite field can be written as the sum of two squares.

This page intentionally left blank

Chapter 14

The Theory of Fields

14.1 Vector Spaces

In order to study fields in depth, we will first need a few results from a first year linear algebra course. However, most linear algebra courses work with vectors and matrices with real numbers for entries, whereas we will generalize the notations to allow arbitrary fields. Nonetheless, most of the proofs will follow the same way for arbitrary fields as for real numbers.

DEFINITION 14.1 Let F be a field. We say that V is a *vector space* over F if V is an abelian group under addition $+$, and for which there is defined a multiplication $a \cdot v$ for all $a \in F$ and $v \in V$ such that:

1. Whenever $a \in F$ and $v \in V$, $a \cdot v \in V$.
2. When $a \in F$, and $v, w \in V$, then $a \cdot (v + w) = a \cdot v + a \cdot w$.
3. When $a, b \in F$, and $v \in V$, then $(a + b) \cdot v = a \cdot v + b \cdot v$.
4. When $a, b \in F$, and $v \in V$, then $(a \cdot b) \cdot v = a \cdot (b \cdot v)$.
5. If e is the identity of F, then $e \cdot v = v$ for all $v \in V$.

The members of V are called *vectors*. The best way to get a feel for vector spaces is to give some examples.

Example 14.1
Consider the set of 3-tuples $\langle u_1, u_2, u_3 \rangle$ where u_1, u_2, and $u_3 \in \mathbb{R}$. Addition of two vectors is done componentwise, and $k \cdot \langle u_1, u_2, u_3 \rangle = \langle ku_1, ku_2, ku_3 \rangle$ when $k \in \mathbb{R}$. This is a vector space over $\mathbb{R}$, and can be denoted by $\mathbb{R}^3$. □

Example 14.2
We can generalize the previous example using any field F in place of $\mathbb{R}$, and consider n-tuples $\langle u_1, u_2, \ldots, u_n \rangle$. Addition is still defined componentwise, and $k \cdot \langle u_1, u_2, \ldots, u_n \rangle = \langle k \cdot u_1, k \cdot u_2, \ldots, k \cdot u_n \rangle$. This will give us a vector space over F, which we can denote by F^n. □

Example 14.3
Let K be a field, and F any subfield of K. Then K is a vector space over F, defining $a \cdot v$ as a product in the field K. Property 5 follows from the fact that the identity of F must also be the identity of K. The other properties follow from the distributive and associative properties of K. □

This last example demonstrates the usefulness in studying vector spaces over a field F. In fact, this is the example that we will concentrate on for the remainder of the chapter.

The next definition is the key to understanding the properties of a vector space.

DEFINITION 14.2 Let V be a vector space over a field F. We say that a finite set $B = \{x_1, x_2, \ldots x_n\}$ of vectors in V are *linearly dependent* if there are elements $c_1, c_2, \ldots c_n \in F$, not all zero, for which

$$c_1x_1 + c_2x_2 + \cdots + c_nx_n = 0.$$

We say that the vectors are *linearly independent* if they are not linearly dependent, that is, if the only way for $c_1x_1 + c_2x_2 + \cdots + c_nx_n = 0$ is for $c_0 = c_1 = \cdots = c_n = 0$.

Example 14.4
The vectors $\langle 1, 4, -1\rangle$, $\langle 2, -3, 1\rangle$, $\langle 4, 5, -1\rangle$ are linearly dependent, since there is a nonzero solution to $c_1\langle 1, 4, -1\rangle + c_2\langle 2, -3, 1\rangle + c_3\langle 4, 5, -1\rangle = 0$, namely $c_1 = 2$, $c_2 = 1$, and $c_3 = -1$. On the other hand, $\langle 2, 0, 1\rangle$, $\langle 0, 0, 3\rangle$, and $\langle 1, 4, 0\rangle$ are linearly independent, since in order to get $c_1\langle 2, 0, 1\rangle + c_2\langle 0, 0, 3\rangle + c_3\langle 1, 4, 0\rangle = 0$, we need $4c_3 = 0$, $2c_1 + c_3 = 0$, and $c_1 + 3c_2 = 0$. This forces $c_3 = 0$, $c_1 = 0$, and $c_2 = 0$, so there are no nonzero solutions. □

DEFINITION 14.3 Let V be a vector space over a field F. A finite set of vectors $\{x_1, x_2, x_3, \ldots x_n\}$ in V is called a *basis of V over F* if the set is linearly independent, and every element of V can be expressed in the form

$$a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n$$

with $a_1, a_2, a_3, \ldots, a_n$ in F.

Here are some examples, all of which are fairly routine to check:

1. The complex numbers $\mathbb{C}$ have a basis $\{1, i\}$ over the real numbers $\mathbb{R}$.
2. The quaternions $\mathbb{H}$ have a basis $\{1, i, j, k\}$ over $\mathbb{R}$.
3. The field $\mathbb{Q}[\sqrt{2}]$ has a basis $\{1, \sqrt{2}\}$ over the rational numbers $\mathbb{Q}$.

4. From example 14.3, the set of real numbers $\mathbb{R}$ is a vector space over the rationals. However, there can be no finite basis $\{x_1, x_2, x_3, \ldots x_n\}$ in $\mathbb{R}$ for which every real number could be expressed as $a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n$, with $a_1, a_2, \ldots a_n \in \mathbb{Q}$, lest the set of reals be countable, which contradicts Cantor's diagonalization theorem (9.1).

There is an easy way to determine if a particular set of vectors is a basis.

LEMMA 14.1
$B = \{x_1, x_2, x_3, \ldots x_n\}$ is a basis of a vector space V over F if, and only if, every element of V can be expressed uniquely *in the form*

$$v = c_1x_1 + c_2x_2 + c_3x_3 + \cdots + c_nx_n.$$

The ordered n-tuple $\langle c_1, c_2, c_3, \ldots, c_n\rangle$ is called the coefficients *of v with respect to B.*

PROOF If B is a basis, then every element $v \in V$ can be expressed in the form $c_1x_1 + c_2x_2 + c_3x_3 + \cdots + c_nx_n$. Suppose that $v = a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n$ is another such expression. Then

$$(a_1 - c_1)x_1 + (a_2 - c_2)x_2 + (a_3 - c_3)x_3 + \cdots + (a_n - c_n)x_n = v - v = 0.$$

But the vectors in B are linearly independent, so the only way that the combination of vectors could be 0 is for $a_i - c_i = 0$ for all $1 \leq i \leq n$. Hence, $a_i = c_i$ for all i, and the representation is unique.

On the other hand, if every $v \in V$ can be uniquely represented as $c_1x_1 + c_2x_2 + c_3x_3 + \cdots + c_nx_n$, then in particular 0 has only one representation, namely $0 = 0x_1 + 0x_2 + 0x_3 + \cdots + 0x_n$. Thus, the vectors in B are linearly independent, and so B is a basis. □

We can define a basis in $\mathbb{Q}^n$ in GAP using the `Basis` command. To find a basis using example 14.4, we enter

```
gap> B := Basis(Rationals^3, [[1,4,-1],[2,-3,1],[4,5,-1]]);
fail
gap> B := Basis(Rationals^3, [[2,0,1],[0,0,3],[1,4,0]]);
Basis( ( Rationals^3 ), [ [ 2, 0, 1 ], [ 0, 0, 3 ],
  [ 1, 4, 0 ] ] )
```

In *Mathematica*®, we use the command **ToBasis**, and enter in "1" for the first argument whenever we are using rational or real numbers for the base field.

B = ToBasis[1, {{1, 4, –1},{2, –3, 1},{4, 5, –1}}]
Error: linearly dependent.
False
B = ToBasis[1, {{2, 0, 1},{0, 0, 3},{1, 4, 0}}]

The first attempt failed because the vectors were linearly dependent. Once we have defined the basis, we can find the coefficients $c_1, c_2, \ldots c_n$ for any element of the vector space.

Coefficients[B,{2,3,4}]

```
gap> Coefficients(B, [2,3,4]);
[ 5/8, 9/8, 3/4 ]
```

This shows that

$$\langle 2,3,4\rangle = \frac{5}{8}\langle 2,0,1\rangle + \frac{9}{8}\langle 0,0,3\rangle + \frac{3}{4}\langle 1,4,0\rangle.$$

LEMMA 14.2
Suppose that V is a vector space over F, and $B = \{x_1, x_2, x_3, \ldots x_n\}$ is a basis of V over F. Then any set $\{y_1, y_2, y_3, \ldots y_n, y_{n+1}\}$ of $n+1$ elements of V is linearly dependent.

PROOF Suppose that $Y = \{y_1, y_2, y_3, \ldots, y_n, y_{n+1}\}$ are linearly independent, so that all of these vectors are nonzero.

Our goal is to show, with a suitable rearrangement of the vectors in B, that $\{y_1, y_2, \ldots y_{k-1}, y_k, x_{k+1}, \ldots, x_n\}$ is a basis for every $0 \leq k \leq n$. If $k = 0$, then this set is the original set B, which is a basis. So let us use induction to assume that it is true for the previous case, that is, that $\{y_1, y_2, \ldots y_{k-1}, x_k, x_{k+1}, \ldots, x_n\}$ is a basis.

We then can express

$$y_k = a_1y_1 + a_2y_2 + \cdots a_{k-1}y_{k-1} + a_kx_k + a_{k+1}x_{k+1} + \cdots + a_nx_n.$$

Since the vectors in Y are linearly independent, we see that at least one of a_k, $a_{k+1} \ldots a_n$ is nonzero. By rearranging the remaining elements of B, we can suppose that $a_k \neq 0$. Then

$$x_k = {a_k}^{-1}(y_k - a_1y_1 - a_2y_2 - \cdots - a_{k-1}y_{k-1} - a_{k+1}x_{k+1} - \cdots - a_nx_n).$$

Any element $v \in V$ can be expressed as $v = c_1y_1 + c_2y_2 + \cdots + c_{k-1}y_{k-1} + c_kx_k + \cdots + c_nx_n$. By substituting for the value of x_k, we see that v can be expressed as a linear combination of $\{y_1, y_2, \ldots, y_{k-1}, y_k, x_{k+1}, \ldots, x_n\}$. If this set were linearly dependent, there would be a nonzero solution to

$$c_1y_1 + c_2y_2 + \cdots + c_{k-1}y_{k-1} + c_ky_k + \cdots + c_nx_n = 0.$$

Then $c_k \neq 0$, lest there also be a nonzero solution to

$$c_1y_1 + c_2y_2 + \cdots + c_{k-1}y_{k-1} + c_kx_k + \cdots + c_nx_n = 0,$$

but we are assuming that $\{y_1, y_2, \ldots y_{k-1}, x_k, x_{k+1}, \ldots, x_n\}$ is a basis. But substituting the value for y_k gives

$$\begin{aligned} c_k \ & (a_1 y_1 + a_2 y_2 + \cdots a_{k-1} y_{k-1} + a_k x_k + \cdots a_n x_n) \\ & + c_1 y_1 + c_2 y_2 + \cdots c_{k-1} y_{k-1} c_{k+1} x_{k+1} + \cdots c_n x_n = 0. \end{aligned}$$

This is a nonzero solution to

$$b_1 y_1 + b_2 y_2 + \cdots + b_{k-1} y_{k-1} + b_k x_k + \cdots b_n x_n = 0,$$

since $b_k = c_k a_k \neq 0$. Thus, the set $\{y_1, y_2, \ldots y_{k-1}, y_k, x_{k+1}, \ldots, x_n\}$ is linearly independent, and hence is a basis of V.

Now we can use the induction to say that $\{y_1, y_2, \ldots, y_n\}$ is a basis of V, but then y_{n+1} can be expressed in terms of $\{y_1, y_2, \ldots, y_n\}$, which shows that Y is in fact linearly dependent. □

We can now use this lemma to show that any two bases must have the same number of elements.

PROPOSITION 14.1

Let V be a vector space over F. If the sets $X = \{x_1, x_2, x_3, \ldots x_n\}$ and $Y = \{y_1, y_2, y_3, \ldots y_m\}$ are both bases of V over F, then $n = m$.

PROOF Suppose that n is not equal to m. By exchanging the roles of X and Y if necessary, we can assume that $n < m$. Then we can use lemma 14.2 to show that $\{y_1, y_2, y_3, \ldots y_{n+1}\}$ is linearly dependent, hence Y is not a basis of V. So we must have $n = m$. □

This proposition allows us to make the following definition.

DEFINITION 14.4 Let V be a vector space over F. If there is a basis $\{x_1, x_2, x_3, \ldots x_n\}$ of V over F, we define the *dimension of V over F* to be the size n of the basis. If there does not exist a finite basis, we say the dimension of V over f is *infinite.*

Looking back at our examples, we see that $\mathbb{R}^3$ is a 3-dimensional vector space over $\mathbb{R}$, $\mathbb{C}$ is a 2-dimensional vector space over $\mathbb{R}$, $\mathbb{H}$ is a 4-dimensional vector space over $\mathbb{R}$, and $\mathbb{R}$ is an infinite-dimensional vector space over $\mathbb{Q}$.

Here is another example. Since Z_3 is a subfield of $GF(9)$, we can view $GF(9)$ as a vector space over Z_3. Let us see if we can find a basis. In *Mathematica* we need to know that the Conway polynomial of degree 2 over Z_3 is $x^2 + 2x + 2$, or $Z(9)^2 = Z(9) + 1$. We can see if $\{1, Z(9)\}$ forms a basis of $GF(9)$ over Z_3.

ConwayPolynomial[3, 2, x]
$2 + 2x + x^2$
InitDomain[3]
Define[Z9^2, Z9 + 1]
B = ToBasis[1, {1, Z9}];
Coefficients[B, Z9^3]
$\{1, 2\}$

```
gap> B := Basis(GF(9), [Z(9)^0,Z(9)]);
Basis( GF(3^2), [ Z(3)^0, Z(3^2) ] )
gap> Coefficients(B, Z(9)^3);
[ Z(3)^0, Z(3) ]
```

This shows that indeed $\{1, Z(9)\}$ is a basis of $GF(9)$ over Z_3, but also that $Z(9)^3 = 1 + Z(3) * Z(9)$. It is logical that $GF(9)$ will be a 2-dimensional vector space over Z_3, since there are 3^2 elements. Likewise, $GF(81)$ is a 4-dimensional vector space over Z_3. But we also can consider $GF(81)$ as a 2-dimensional vector space over $GF(9)$. In GAP we can use `AsVectorSpace` so that GAP will view $GF(81)$ as a vector space over $GF(9)$ instead of the natural $GF(3)$.

```
gap> V := AsVectorSpace(GF(9),GF(81));
AsField( GF(3^2), GF(3^4) )
gap> B := Basis(V, [Z(81),Z(3)]);
Basis( AsField( GF(3^2), GF(3^4) ), [ Z(3^4), Z(3) ] )
gap> Coefficients(B,Z(81)^2);
[ Z(3^2)^3, Z(3^2) ]
```

This shows that $Z(81)$ and $Z(3)$ form a basis of $GF(81)$ over $GF(9)$, for example, the element $Z(81)^2$ can be expressed as $Z(9)^3 \cdot Z(81) + Z(9) \cdot Z(3)$.

To do this in *Mathematica*, we must enter in a basis for the root vector space as the first argument for the **ToBasis** command. Also, we have to define $Z(81)$ in *Mathematica* so that the original $Z(9)$ will generate a subfield. We can borrow the result from GAP, that

$$Z(81)^2 = Z(9)^3 \cdot Z(81) + Z(9) \cdot Z(3) = (1 + 2Z(9)) \cdot Z(81) + 2Z(9).$$

Define[Z81^2, (1 + 2 Z9)*Z81 + 2 Z9]
B = ToBasis[{1, Z9}, {Z81, 2}];
Coefficients[B, Z81^2]
$\{1 + 2\,\text{Z9}, \text{Z9}\}$

This last example shows that it is possible to have a vector space over a vector space, if the later vector space happens to be a field. What can we say about the dimension of a vector space over a vector space?

PROPOSITION 14.2

If E is a vector space over F of dimension m, which also happens to be a field, and V is a vector space over E of dimension n, then V is a vector space

of F of dimension $m \cdot n$. Furthermore, if $\{x_1, x_2, x_3, \ldots x_m\}$ is a basis of E over F, and $\{y_1, y_2, y_3, \ldots y_n\}$ is a basis of V over E, then the set

$$\begin{aligned} S = \{\ & x_1y_1,\ x_2y_1, x_3y_1, \ldots x_my_1, \\ & x_1y_2,\ x_2y_2, x_3y_2, \ldots x_my_2, \\ & x_1y_3,\ x_2y_3, x_3y_3, \ldots x_my_3, \\ & \qquad\qquad \ldots\ldots\ldots \\ & x_1y_n,\ x_2y_n, x_3y_n, \ldots x_my_n\} \end{aligned}$$

is a basis of V over F.

PROOF Since $\{y_1, y_2, y_3, \ldots, y_n\}$ is a basis for V over E, we can write any element of V in the form

$$c_1y_1 + c_2y_2 + c_3y_3 + \cdots + c_ny_n,$$

where $c_1, c_2, c_3, \ldots, c_n$ are in E.

Since $\{x_1, x_2, x_3, \ldots x_m\}$ is a basis of E over F, we can in turn write

$$\begin{aligned} c_1 &= a_{1,1}x_1 + a_{2,1}x_2 + a_{3,1}x_3 + \cdots a_{m,1}x_m, \\ c_2 &= a_{1,2}x_1 + a_{2,2}x_2 + a_{3,2}x_3 + \cdots a_{m,2}x_m, \\ c_3 &= a_{1,3}x_1 + a_{2,3}x_2 + a_{3,3}x_3 + \cdots a_{m,3}x_m, \\ &\qquad \ldots\ldots\ldots \\ c_n &= a_{1,n}x_1 + a_{2,n}x_2 + a_{3,n}x_3 + \cdots a_{m,n}x_m, \end{aligned}$$

where each $a_{i,j}$ is in F. Combining these, we see that every element of E can be expressed in the form

$$\begin{aligned} a_{1,1}x_1y_1 \ &+\ a_{2,1}x_2y_1 + a_{3,1}x_3y_1 + \cdots + a_{m,1}x_my_1 \\ + a_{1,2}x_1y_2 \ &+\ a_{2,2}x_2y_2 + a_{3,2}x_3y_2 + \cdots + a_{m,2}x_my_2 \\ + a_{1,3}x_1y_3 \ &+\ a_{2,3}x_2y_3 + a_{3,3}x_3y_3 + \cdots + a_{m,3}x_my_3 \\ &\qquad \ldots\ldots\ldots \\ + a_{1,n}x_1y_n \ &+\ a_{2,n}x_2y_n + a_{3,n}x_3y_n + \cdots + a_{m,n}x_my_n. \end{aligned}$$

Thus, to show that the set S is a basis of V over F, we merely have to show that these vectors are linearly independent. Let us switch to a summation notation for the remainder of the proof. Suppose that there is a nonzero linear combination of these vectors that produces 0, that is

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}x_iy_j = 0$$

for $a_{i,j}$ in F. Then we have

$$0 = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j}x_iy_j = \sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_{i,j}x_i\right)y_j.$$

Since $\{y_1, y_2, y_3, \ldots, y_n\}$ is a basis of V over E, the only way that the right hand expression could be zero is if

$$\sum_{i=1}^{m} a_{i,j} x_i = 0$$

for all $j = 1, 2, 3, \ldots n$. Now $\{x_1, x_2, x_3, \ldots x_m\}$ is a basis of E over F, so the only way that each of these sums could be 0 is if $a_{i,j} = 0$ for all values of i and j. Since all of the coefficients must be 0, the vectors in S are linearly independent, and therefore the S is a basis of V over F of dimension $m \cdot n$. □

The main use of vector spaces in abstract algebra is in the case where the vector space happens to be a field. We will explore this possibility in the next section.

14.2 Extension Fields

In the last section, we found that many of the examples of vector spaces turned out to also be fields. We will give a special name to this situation.

DEFINITION 14.5 If F is a nontrivial subfield of K, and K is a finite-dimensional vector space over F, we say that K is a *finite extension* of F. We say the degree, or *dimension* of the extension is the size of a basis $\{x_1, x_2, x_3, \ldots x_n\}$ of K over F.

For example, the complex numbers $\mathbb{C}$ are a 2-dimensional extension of $\mathbb{R}$. The quaternions $\mathbb{H}$ are a 4-dimensional extension of $\mathbb{R}$. The field $GF(27)$ is a 3-dimensional extension of Z_3, regardless of which basis we use.

It seems intuitively obvious that isomorphic fields have the same dimension over some field F contained in both of the fields. Yet this is only true if the isomorphism ϕ maps the base field F to itself.

PROPOSITION 14.3
If K and E are two finite extensions of F, and suppose that there is an isomorphism ϕ from K onto E such that $\phi(x) = x$ for all x in F, then K and E have the same dimension over F.

PROOF Suppose that $\{x_1, x_2, x_3, \ldots x_n\}$ is a basis of K over F. We want to show that $\{\phi(x_1), \phi(x_2), \phi(x_3), \ldots, \phi(x_n)\}$ is a basis of E over F. If v is in E, then $\phi(u) = v$ for some u in K. Since K is generated by the elements in

the basis, we have

$$u = c_1x_1 + c_2x_2 + c_3x_3 + \cdots c_nx_n$$

for some $c_1, c_2, c_3, \ldots, c_n$ in F. Then

$$\begin{aligned} v = \phi(u) &= \phi(c_1)\phi(x_1) + \phi(c_2)\phi(x_2) + \phi(c_3)\phi(x_3) + \cdots + \phi(c_n)\phi(x_n) \\ &= c_1\phi(x_1) + c_2\phi(x_2) + c_3\phi(x_3) + \cdots + c_n\phi(x_n). \end{aligned}$$

Thus, $\{\phi(x_1), \phi(x_2), \phi(x_3), \ldots, \phi(x_n)\}$ generates the field E. Also, if

$$c_1\phi(x_1) + c_2\phi(x_2) + c_3\phi(x_3) + \cdots + c_n\phi(x_n) = 0,$$

then $\phi(c_1x_1 + c_2x_2 + c_3x_3 + \cdots c_nx_n) = 0$, which implies that

$$c_1x_1 + c_2x_2 + c_3x_3 + \cdots c_nx_n = 0$$

since K and E are isomorphic. But since $\{x_1, x_2, x_3, \ldots x_n\}$ is a basis for K, this can only happen if $c_1 = c_2 = c_3 = \cdots c_n = 0$. So

$$\{\phi(x_1), \phi(x_2), \phi(x_3), \ldots, \phi(x_n)\}$$

is a basis for E over F, and hence K and E have the same dimension over the field F. □

If K is a finite extension of a field F, then F is a subfield of K. Of course there will probably be many other subfields of K, and we need a way to identify these subfields. We have already seen how to find the smallest subgroup or a subring that contains certain elements, and we can follow the same logic for subfields.

DEFINITION 14.6 Let K be a field, and let E be a field containing the field K. Let S be a set of elements in E. Let L denote the collection of all subfields of E that contain the field K, along with the set S. Then we define

$$K(S) = \bigcap_{H \in L} H.$$

That is, $K(S)$ is the intersection of all subfields of E that contain both K and S. If $S = a_1, a_2, a_3, \ldots a_n$, we will write $K(a_1, a_2, a_3, \ldots a_n)$ for $K(S)$. Thus, if S consists of a single element a, we can write $K(a)$ for $K(S)$.

LEMMA 14.3
Let K be a subfield of E, and let S be a collection of elements of E. Then $K(S)$ is the smallest field that contains both K and the elements S.

PROOF First, we must show that $K(S)$ is a subfield of E. If x and y are in $K(S)$, $y \neq 0$, then x and y are in each of the subfields in the collection L. Then $x - y$ and $x \cdot y^{-1}$ are also in each of the subfields in this collection. Thus, $x - y$ and $x \cdot y^{-1}$ are in $K(S)$, and so $K(S)$ is a subfield of E.

To show that $K(S)$ is the smallest field containing both K and the elements S, note that $K(S)$ is one of the subfields in the collection L. Thus, any subfield containing K and the elements of S must also contain $K(S)$. □

For example, If K is the real numbers, and $i = \sqrt{-1}$, then $\mathbb{R}(i)$ gives us the complex numbers $\mathbb{C}$. The field $\mathbb{Q}(\sqrt{2})$ is the smallest field containing $\mathbb{Q}$ and $\sqrt{2}$, which happens to be the same as the ring $\mathbb{Q}[\sqrt{2}]$.

The strategy for defining a field extension in GAP or *Mathematica* is very similar to that of defining a finite field. We begin by finding an irreducible polynomial $f(x)$ in the field F, and creating the field $K = F[x]/(f(x))$.

PROPOSITION 14.4

Let F be a field, and let $f(x)$ be an irreducible polynomial in $F[x]$ of degree d. Then the field $K = F[x]/(f(x))$ is a finite extension of F of dimension d.

PROOF From proposition 13.1, $K = F[x]/(f(x))$ is a field that contains F as a subfield. Let $y = x + (f(x))$ in K. If we treat $f(x)$ as a polynomial in $K[x]$, we find that $f(y) = 0$. Consider the set $\{1, y, y^2, y^3, \cdots y^{n-1}\}$. We wish to show that this set is a basis for K. That is, we wish to show that every element of K can be expressed uniquely as

$$k = a_1 1 + a_2 y + a_3 y^2 + \cdots + a_n y^{n-1},$$

where the $a_1, a_2, a_3, \ldots, a_n$ are in F. Any element $k \in K$ can be expressed as $k = g(x) + (f(x))$ for some polynomial $g(x)$ in $F[x]$. By the division algorithm theorem (12.1), there exist unique polynomials $q(x)$ and $r(x)$ such that

$$g(x) = f(x) \cdot q(x) + r(x),$$

where either $r(x) = 0$, or the degree of $r(x)$ is less than n. Then

$$r(x) = a_1 + a_2 x + a_3 x^2 + \cdots + a_n x^{n-1}$$

for some $a_1, a_2, a_3, \ldots, a_n$ in F. Note that we can now write

$$k = g(x) + (f(x)) = r(x) + (f(x)) = a_1 + a_2 y + a_3 y^2 + \cdots + a_n y^{n-1}.$$

Since $r(x)$ is unique, k is uniquely determined as a linear combination of $\{1, y, y^2, \ldots, y^{n-1}\}$. Thus, by lemma 14.1, $\{1, y, y^2, \ldots, y^{n-1}\}$ is a basis. □

Let us look at an example in *Mathematica*. Let F be the field of rational numbers, and let $f(x) = x^3 - 2$. Since the characteristic of $\mathbb{Q}$ is 0, we begin the definition by the command

InitDomain[0]

Next, we let a be a root to the equation $x^3 - 2$. That is, we define a^3 to be 2.

Define[a^3, 2]

That's all there is to it! The basis of this extension field is $\{1, a, a^2\}$. We can verify this with *Mathematica*.

CheckField[{1, a, a^2}]

CheckField actually does more than just verify that the ring is a field. It also allows us to do divisions in this field.

1/(a + a^2)

will compute $1/(\sqrt[3]{2} + \sqrt[3]{4}) = (2\sqrt[3]{2} + \sqrt[3]{4} - 2)/6$.

In GAP, we must first define x to be an indeterminate over the rationals, so that we can express the polynomial $x^3 - 2$ in $\mathbb{Q}[x]$. We then can use the command `FieldExtension` to create the extension field.

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals,x^3-2);
<algebraic extension over the Rationals of degree 3>
gap> a := PrimitiveElement(A);
a
gap> 1/(a+a^2);
1/6*a^2+1/3*a-1/3
```

This introduces the command `PrimitiveElement`, which defines the letter a to be the element of the field for which $a^3 = 2$. We see that GAP is already able to compute divisions in this new field. We can verify that $\{1, a, a^2\}$ is a basis.

```
gap> B := Basis(A,[a^0,a,a^2]);
Basis( <algebraic extension over the Rationals of degree 3>,
  [ !1, a, a^2 ] )
```

Note that GAP writes `!1` for the identity element of this new field. This distinguishes it from the rational number 1. However, you do not enter `!1` into GAP, but rather `a^0`.

Although this example demonstrates that any extension field of the form $F[x]/(f(x))$ can be entered into GAP or *Mathematica*, we would like to show that *any* extension field can be entered into *Mathematica* or GAP in the same way. That is, we must show that any finite extension of F is isomorphic to $F[x]/(f(x))$ for some polynomial $f(x)$.

PROPOSITION 14.5

Suppose a field K is a finite extension of F of dimension n. Let y be an element of K. Then there is an irreducible polynomial $f(x)$ in $F[x]$ of degree

at most n such that $f(y) = 0$. That is, when $f(x)$ is treated as a polynomial in $K[x]$, y is a root of $f(x)$. Furthermore, there is a unique polynomial of lowest degree that satisfies these conditions and for which the leading coefficient is equal to 1.

PROOF Consider the set $\{1, y, y^2, y^3, \ldots, y^n\}$. Since there are $n + 1$ elements in this set, and K has dimension n over F, by lemma 14.2 these are linearly dependent, so there is a nonzero solution to

$$a_0 + a_1 y + a_2 y^2 + a_3 y^3 + \cdots + a_n y^n = 0$$

with $a_0, a_1, a_2, \cdots, a_n$ in F. Thus, there is a nonzero polynomial

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n$$

in $F[x]$ for which y is a root when treated as a polynomial in $K[x]$.

Let us now show uniqueness. Let $f(x)$ be a polynomial of lowest possible degree in $F[x]$ such that $f(y) = 0$. Since F is a field, we can divide this polynomial by its leading coefficient to obtain a polynomial with a leading coefficient of 1. Now, if there were two such polynomials, $f(x)$ and $g(x)$, then by the division algorithm theorem (12.1), there exist polynomials $q(x)$ and $r(x)$ such that $f(x) = g(x) \cdot q(x) + r(x)$, where either $r(x) = 0$ or the degree of $r(x)$ is strictly less than the degree of $g(x)$. But note that

$$0 = f(y) = g(y) \cdot q(y) + r(y) = 0 + r(y) = 0.$$

Thus, y is a root of the polynomial $r(x)$. But the degree of $f(x)$ and $g(x)$ was chosen to be minimal. So $r(x) = 0$, and $f(x)$ is a multiple of $g(x)$. Finally, since both $f(x)$ and $g(x)$ have the same degree and have the same leading term of 1, we have $f(x) = g(x)$. Therefore, there is a unique polynomial in $F[x]$ of minimal degree and leading coefficient of 1 such that $f(y) = 0$. ∎

The unique polynomial in proposition 14.5 will be given a special name.

DEFINITION 14.7 If a field K is a finite extension of F, and a is an element of K, we define the polynomial $f(x)$ given by proposition 14.5 that has a leading coefficient of 1 to be the *irreducible polynomial of a over F,* denoted $\text{Irr}_F(a, x)$.

For example, $\text{Irr}_{\mathbb{Q}}(\sqrt{2}, x) = x^2 - 2$, since $x^2 - 2$ is the simplest polynomial with rational coefficients for which $\sqrt{2}$ is a root. Note that if we were to allow real coefficients, we could come up with a simpler polynomial: $\text{Irr}_{\mathbb{R}}(\sqrt{2}, x) = x - \sqrt{2}$. Finally, consider the number $\cos(\pi/9)$. We found in section 11.4 that this number is a root of the polynomial $4x^3 - 3x - \frac{1}{2}$. However, we want the leading coefficient of the polynomial to be 1, so we write

$$\text{Irr}_{\mathbb{Q}}(\cos(\pi/9), x) = x^3 - \frac{3x}{4} - \frac{1}{8}.$$

Once we find the irreducible polynomial for an element a, it is not hard to program *Mathematica* or GAP to mimic the field $\mathbb{Q}(a)$. For example, let us enter the field $\mathbb{Q}(\cos(\pi/9))$ into *Mathematica*. If we let $a = \cos(\pi/9)$, we can enter the field by the commands

InitDomain[0]
Define[a^3, 3 a/4 + 1/8]

The first command tells *Mathematica* that we are working with a field of characteristic 0, and the second command identifies a as one solution to the equation $x^3 - 3x/4 - 1/8$. We can check that this is a field with the *Mathematica* command

CheckField[{1, a, a^2}]

which will also allow division operations to be performed in this field.

The corresponding commands in GAP are

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals,x^3-3*x/4-1/8);
<algebraic extension over the Rationals of degree 3>
gap> a := PrimitiveElement(A);
a
```

Have we really defined the field $\mathbb{Q}(\cos(\pi/9))$? Actually, we have defined the field

$$\mathbb{Q}[x]/(x^3 - 3x/4 - 1/8)$$

in GAP or *Mathematica*, but we can prove that these two fields are isomorphic.

PROPOSITION 14.6
Let F be a subfield of K, and suppose $f(x)$ is an irreducible polynomial in $F[x]$ that has a root w in the larger field K. Then

$$F(w) \approx F[x]/(f(x)).$$

PROOF Let us consider the evaluation homomorphism ϕ_w that maps polynomials in $F[x]$ to elements in $F(w)$:

$$\phi_w(g(x)) = g(w).$$

By proposition 12.1, ϕ_w is a ring homomorphism. The image of this homomorphism contains both F and w, and since $F(w)$ is the smallest field containing both F and w, the image is all of $F(w)$. The kernel of ϕ_w is the set of polynomials in $F[x]$ that have w as a root. But $f(x)$ is an irreducible polynomial in $F[x]$ containing w as a root. Thus, any polynomial in the kernel is a multiple of $f(x)$. Thus, the kernel of ϕ_w is $(f(x))$. Finally, by the first ring isomorphism theorem (10.2), we have that $F(w) \approx F[x]/(f(x))$. □

It is now easy to see that the dimension of the field extension $F(u)$ will be the dimension of the irreducible polynomial $f(x) = \text{Irr}_F(u, x)$.

COROLLARY 14.1
Let K be a finite extension of a field F, and let u be an element in K. If $f(x) = \text{Irr}_F(u, x)$ has degree n, then $F(u)$ has dimension n over F.

PROOF By proposition 14.5, $f(x) = \text{Irr}_F(u, x)$ exists. By proposition 14.6, $F(u)$ is isomorphic to the field $F[x]/(f(x))$. By proposition 14.4, $F[x]/(f(x))$ has dimension n over F. Finally, by corollary 14.3, two isomorphic extensions of F must have the same dimension over F provided that the isomorphism fixes the elements of F, which the isomorphism in proposition 14.6 clearly does. Thus, the dimension of $f(u)$ over F is n. □

Notice that we never had to tell *Mathematica* or GAP that $a = \cos(\pi/9)$ in our definition of $\mathbb{Q}(\cos(\pi/9))$. Rather, we merely entered the information that a satisfies the equation $a^3 - 3a/4 - 1/8 = 0$.

But there are two *other* solutions to this equation, namely $-\cos(2\pi/9)$ and $\cos(4\pi/9)$. How does *Mathematica* or GAP know that the field is not $\mathbb{Q}(-\cos(2\pi/9))$ or $\mathbb{Q}(\cos(4\pi/9))$?

The answer is of course that these fields are both isomorphic to $\mathbb{Q}(\cos(\pi/9))$, so *Mathematica* or GAP didn't need to know the exact value of a. In fact, we can prove that if we start with isomorphic fields, and extend both of them by two elements for which the irreducible polynomials correspond, then the two field extensions will be isomorphic.

PROPOSITION 14.7
Let f be an isomorphism between a field K and a field E. Let M be a finite extension of K, and let u be in M. Let

$$p(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots + c_nx^n$$

be $\text{Irr}_K(u, x)$. Define

$$h(x) = f(c_0) + f(c_1)x + f(c_2)x^2 + f(c_3)x^3 + \cdots + f(c_n)x^n$$

which is in $E[x]$. Suppose there is a finite extension of E for which there is a root of $h(x)$, called v. Then there is an isomorphism μ from $K(u)$ to $E(v)$ for which $\mu(u) = v$, and $\mu(x) = f(x)$ for all x in K.

PROOF By lemma 13.3, we can extend f to a isomorphism from $K[x]$ to $E[x]$. By proposition 12.1, ϕ_v is a ring homomorphism from $E[x]$ to $E(v)$. We can combine these homomorphisms to produce the homomorphism

$$f \cdot \phi_v : K[x] \to E[x] \to E(v).$$

Since the isomorphism in lemma 13.3 sends x to x, we have that $(f \cdot \phi_v)(x) = \phi_v(f(x)) = \phi_v(x) = v$. So v is in the image of this combination of homomorphisms, as well as the subfield E. Thus, the image of $f \cdot \phi_v$ is $E(v)$. The kernel of ϕ_v is the set of polynomials in $E[x]$ with v as a root. But $h(x)$ is an irreducible polynomial in $E[x]$ for which $h(v) = 0$. Thus, the kernel of ϕ_v is the ideal $(h(x))$. Since $h(x) = f(p(x))$, we have that the kernel of $f \cdot \phi_v$ is $(p(x))$. Thus, by the first ring isomorphism theorem (10.2),

$$K[x]/(p(x)) \approx E(v).$$

By proposition 14.6, we also have

$$K(u) \approx K[x]/(p(x)),$$

and in this isomorphism, u mapped to the coset $x + (p(x))$. If we let μ be the combination of these two isomorphisms,

$$\mu : K(u) \to K[x]/(p(x)) \to E(v),$$

then $\mu(u) = \phi_v(f(x)) = v$, and $\mu(x) = f(x)$ for all x in K. □

The usual application of this proposition is when K and E are the same field, as in the case $\mathbb{Q}(\cos(\pi/9))$ and $\mathbb{Q}(-\cos(2\pi/9))$, in which case we not only can prove that $\mathbb{Q}(\cos(\pi/9))$ and $\mathbb{Q}(-\cos(2\pi/9))$ are isomorphic, but we can impose further conditions on the isomorphism.

COROLLARY 14.2

If K is a finite extension of a field F, and u and v are two elements in K such that $\mathrm{Irr}_F(u, x) = \mathrm{Irr}_F(v, x)$, then there is an isomorphism μ between $F(u)$ and $F(v)$ such that $\mu(u) = v$, and $\mu(x) = x$ for all x in F.

PROOF We simply let f be the identity mapping from F to itself, and use proposition 14.7. Then $p(x)$ and $h(x)$ are both equal to $\mathrm{Irr}_F(u, x)$. Since v is another root of $h(x)$ the conclusion follows from the conclusion of proposition 14.7. □

We discovered in section 13.2 that every finite field could be expressed in the form $Z_p[x]/(f(x))$, with $f(x)$ an irreducible polynomial in $Z_p[x]$. It is natural to ask whether any finite extension of a field can be represented in the form $F[x]/(f(x))$ for some polynomial $f(x)$ in $F[x]$. Although there are some fields that are exceptions, $\mathbb{Q}$ and $\mathbb{R}$ are not among them. Once we have proven this, we will be able to enter any finite extension of $\mathbb{Q}$ or $\mathbb{R}$ into *Mathematica* using the same technique that was used for finite fields.

14.3 Splitting Fields

We have already seen that given an irreducible polynomial $f(x)$ in $F[x]$, we can construct a field $F[x]/(f(x))$ for which $f(x)$ has a root in this new field. This raises an interesting question: Can we construct a field for which $f(x)$ factors *completely* in the new field? Let us demonstrate with some examples. Let $f(x) = x^3 + x^2 - 2x - 1$. We begin by showing that this polynomial is irreducible over the rationals.

Factor[x^3 + x^2 − 2 x − 1]

Unless otherwise specified, *Mathematica* will factor polynomials over the field $\mathbb{Q}$.

In GAP, we must first declare x to be a variable over the rationals

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Factor(x^3+x^2-2*x-1,Rationals);
[ x^3+x^2-2*x-1 ]
```

Since the output is essentially unchanged, this indicates that the polynomial is irreducible.

If a is one root of this polynomial, we can define $\mathbb{Q}(a)$ in *Mathematica* as follows, and find the factorization by including the a as a second parameter of the **Factor** command.

InitDomain[0]
Define[a^3, − a^2 + 2 a + 1]
Factor[x^3 + x^2 − 2 x − 1, a]

In GAP, we have to do a few more steps. First we define the extension field over this polynomial.

```
gap> A := FieldExtension(Rationals,x^3+x^2-2*x-1);
<algebraic extension over the Rationals of degree 3>
```

In order to factor the polynomial over the new field, we must first declare a new variable y to be an indeterminate of this new field.

```
gap> y := Indeterminate(A,"y");
y
gap> Factor(y^3+y^2-2*y-1,A);
[ y+(-a), y+(-a^2+2), y+(a^2+a-1) ]
```

This shows that the polynomial $x^3 + x^2 - 2x - 1$ factors completely as

$$(x - a)(x - a^2 + 2)(x + a^2 + a - 1)$$

in the field $\mathbb{Q}(a)$. Notice that GAP automatically displays the root of the polynomial as a, even though we never specified this. In fact, GAP always

uses a to display the primitive element, but to enter an expression involving a, we must first set a to the primitive element.

In this case, creating an extension field allowed the polynomial to factor completely in the new field. In fact, this is very similar to what we discovered for finite fields. However, this will not always be the case. Consider the irreducible polynomial $x^3 - 2$. The factorization of this polynomial in $\mathbb{Q}(\sqrt[3]{2})$ is

$$(x - \sqrt[3]{2})(x^2 + \sqrt[3]{2}x + \sqrt[3]{4}).$$

Since the other two roots are complex, the quadratic term must be irreducible over $\mathbb{Q}(\sqrt[3]{2})$, since it is irreducible over the real numbers.

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals,x^3-2);
<algebraic extension over the Rationals of degree 3>
gap> x := Indeterminate(A,"x");
x
gap> Factor(x^3-2,A);
[ x+(-a), x^2+a*x+a^2 ]
```

Here, we reused the variable x, even though this overwrites the original x, since we no longer need the original x.

In *Mathematica*, this factorization can be found by entering

InitDomain[0]
Define[a^3, 2]
Factor[x^3 − 2, a]
$(-a + x)(a^2 + ax + x^2)$

How can we get the polynomial $x^3 - 2$ to factor completely into linear terms? We can define a new element, b, to be a root of the irreducible quadratic. That is, we use the "extension of an extension" $\mathbb{Q}(\sqrt[3]{2}, b)$, where b satisfies $a^2 + ab + b^2 = 0$, that is, $b^2 = -\sqrt[3]{4} - b\sqrt[3]{2}$.

Define[b^2, −a^2 − a b]
Factor[x^3 − 2, a, b]
$(-a + x)(-b + x)(a + b + x)$

```
gap> a := PrimitiveElement(A);
a
gap> B := FieldExtension(A, x^2 + a*x + a^2);
<algebraic extension over the Rationals of degree 6>
gap> x := Indeterminate(B, "x");
x
gap> Factor(x^3 - 2,B);
[ x+(!-a), x+(-a), x+(a+a) ]
```

Notice that $\mathbb{Q}(\sqrt[3]{2})$ is a 3-dimensional extension of $\mathbb{Q}$, and $\mathbb{Q}(\sqrt[3]{2}, b)$ is a 2-dimensional extension of $\mathbb{Q}(\sqrt[3]{2})$. Thus, by proposition 14.2, $\mathbb{Q}(\sqrt[3]{2}, b)$ is a 6-dimensional extension of $\mathbb{Q}$.

Since GAP always displays the primitive element as `a`, this gets a little confusing when we have an extension of an extension. Sometimes the distinction is shown with an extra ! sign, but not always. In this case we can figure out logically that the factors must be $(x-a)(x-b)(x+a+b)$, but a much easier way is to use the command

```
gap> ViewFactors(last,B,["a","b"]);
[ x-a, x-b, x+a+b ]
```

The middle argument is the field that the factors are in, and the list of strings shows how the primitive elements are to be displayed.

A longer example of this process is the polynomial $x^4 - x + 1$.

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Factor(x^4-x+1,Rationals);
[ x^4-x+1 ]
gap> A := FieldExtension(Rationals,x^4-x+1);
<algebraic extension over the Rationals of degree 4>
gap> x := Indeterminate(A,"x");
x
gap> Factor(x^4-x+1,A);
[ x+(-a), x^3+a*x^2+a^2*x+(a^3-1) ]
gap> a := PrimitiveElement(A);
a
gap> B := FieldExtension(A, x^3 + a*x^2 + a^2*x + a^3 - 1);
<algebraic extension over the Rationals of degree 12>
gap> x := Indeterminate(B,"x");
x
gap> Factor(x^4-x+1,B);
[ x+(!-a), x+(-a), x^2+(a+a)*x+(a^2+a*a+a^2)]
gap> ViewFactors(last,B,["a","b"]);
[ x-a, x-b, x^2+x*a+x*b+a^2+a*b+b^2 ]
gap> b := PrimitiveElement(B);
a
gap> a := a*One(b);
!a
gap> C := FieldExtension(B, x^2+x*a+x*b+a^2+a*b+b^2 );
<algebraic extension over the Rationals of degree 24>
gap> x := Indeterminate(C,"x");
x
gap> Factor(x^4-x+1,C);
[ x+(!!-a), x+(!-a)), x+(-a), x+(a+(a+a)) ]
gap> ViewFactors(last,C,["a","b","c"]);
[ x-a, x-b, x-c, x+a+b+c ]
```

GAP has a problem multiplying the primitive element of A with an element of C, since C is not a direct extension of A. This is why we had to replace a with `a*One(b)`, which is the corresponding element of B.

Not only is the polynomial irreducible, but each time we create an extension in *Mathematica* or GAP that forces another root to the equation, the remaining polynomial refuses to factor in the new field extension. Thus, it requires three field extensions before it finally factors completely. By this time, the final extension is a 24 dimensional over the rational numbers $\mathbb{Q}$. Yet from this

example it is easy to see that this procedure could be carried out over any polynomial.

LEMMA 14.4
Let F be a field, and let $f(x)$ be a polynomial in $F[x]$ of degree n whose leading coefficient is c_n. Then there is a finite extension K of F such that

$$f(x) = c_n \cdot (x - u_1) \cdot (x - u_2) \cdot (x - u_3) \cdots (x - u_n),$$

where $u_1, u_2, u_3, \ldots u_n$ are elements in K. Furthermore, the dimension of K over F is at most $n!$.

PROOF The proof is by induction on n. If $n = 1$, then $f(x)$ is a linear function, so its only root is in F. Thus $K = F$, and the degree of K over F is $1 = 1!$.

Suppose that this is true for polynomials of degree less than n. Let $p(x)$ be an irreducible factor of $f(x)$, and consider the field $E = F[x]/(p(x))$. By proposition 14.4, E is a finite extension of F whose dimension over F is the degree of $p(x)$, which is at most n. Then $u_n = x + (p(x))$ is a root of $p(x)$ in the field E, and since $p(x)$ is a factor of $f(x)$, $(x - u_n)$ is a factor of $f(x)$ in the field E. Thus, we can write $f(x) = g(x) \cdot (x - u_n)$ for some $g(x)$ in $E[x]$. Note that $g(x)$ has degree $(n - 1)$, and has the same leading coefficient as $f(x)$. Thus, we can use the induction hypothesis to show that there is a field K that is a finite extension of E with dimension at most $(n - 1)!$ such that $g(x)$ factors completely as

$$g(x) = c_n \cdot (x - u_1) \cdot (x - u_2) \cdot (x - u_3) \cdot \cdots \cdot (x - u_{n-1}).$$

Thus,

$$f(x) = c_n \cdot (x - u_1) \cdot (x - u_2) \cdot (x - u_3) \cdot \cdots \cdot (x - u_{n-1}) \cdot (x - u_n).$$

By proposition 14.2, the dimension of K over F is the product of the dimension of E over F times the dimension of K over E. Thus, the dimension of K over F is at most $n \cdot (n - 1)! = n!$. □

DEFINITION 14.8 If K is a field for which the polynomial $f(x)$ in $F[x]$ factors as

$$f(x) = c_n \cdot (x - u_1) \cdot (x - u_2) \cdot (x - u_3) \cdots (x - u_n),$$

then the field $F(u_1, u_2, u_3, \ldots u_n)$ is called the *splitting field* for the polynomial $f(x)$.

For example, the splitting field of x^3+x^2-2x-1 was found to be $\mathbb{Q}(a)$, where a is one root of the polynomial. Thus, the splitting field is a 3-dimensional

extension of $\mathbb{Q}$. The splitting field of $x^3 - 2$ turned out to be a 6-dimensional extension of $\mathbb{Q}$. The splitting field of $x^4 - x + 1$ turned out to be a 24-dimensional extension of $\mathbb{Q}$. Lemma 14.4 points out that this is the largest possible dimension of a fourth degree polynomial.

The splitting field for the polynomial $x^5 - 5x + 12$ turns out to be rather interesting. When we factor this over the field $\mathbb{Q}(a)$, where a is a root of the polynomial,

InitDomain[0]
Define[a^5, 5 a – 12]
Factor[x^5 – 5 x + 12, a]

$$(-a+x)\left(2-\frac{5a}{4}-\frac{a^2}{4}-\frac{a^3}{4}-\frac{a^4}{4}+x+\frac{3ax}{4}-\frac{a^2x}{4}-\frac{a^3x}{4}-\frac{a^4x}{4}+x^2\right)$$
$$\left(-1-\frac{a}{2}-\frac{a^3}{2}-x+\frac{ax}{4}+\frac{a^2x}{4}+\frac{a^3x}{4}+\frac{a^4x}{4}+x^2\right)$$

we find it doesn't split completely. We can let b be a root to the last polynomial, and try again.

Define[b^2, 1 + a/2 + a^3/2 + b – (a + a^2 + a^3 + a^4) b/4]
Factor[x^5 – 5 x + 12, a,b]

$$(-a+x)\,(-b+x)\left(-1+\frac{a}{4}+\frac{a^2}{4}+\frac{a^3}{4}+\frac{a^4}{4}+b+x\right)$$
$$\left(\frac{3}{2}+\frac{a}{4}-\frac{a^2}{4}-\frac{a^3}{4}-\frac{a^4}{4}-\frac{b}{2}-\frac{ab}{2}+x\right)\left(\frac{-1}{2}+\frac{a}{2}+\frac{b}{2}+\frac{ab}{2}+x\right)$$

This time, the polynomial factors completely in $\mathbb{Q}(a, b)$. Hence the splitting field is 10-dimensional over $\mathbb{Q}$. To do this in GAP, we enter:

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Factor(x^5-5*x+12,Rationals);
[ x^5-5*x+12 ]
gap> A := FieldExtension(Rationals,x^5-5*x+12);
<algebraic extension over the Rationals of degree 5>
gap> x := Indeterminate(A,"x");
x
gap> Factor(x^5-5*x+12,A);
[ x+(-a),
  x^2+(-1/4*a^4-1/4*a^3-1/4*a^2+3/4*a+1)*x+
    (-1/4*a^4-1/4*a^3-1/4*a^2-5/4*a+2),
  x^2+(1/4*a^4+1/4*a^3+1/4*a^2+1/4*a-1)*x+(-1/2*a^3-1/2*a-1) ]
gap> 4*ViewFactors(last,A,["a"]);
[ !4*x+(!-4)*a,
  -x*a^4-x*a^3-a^4-x*a^2-a^3+!4*x^2+!3*x*a-a^2+!4*x+(!-5)*a+!8,
  x*a^4+x*a^3+x*a^2+(!-2)*a^3+!4*x^2+x*a+(!-4)*x+(!-2)*a+(!-4) ]
gap> a := PrimitiveElement(A);
a
gap> B:=FieldExtension(A,x^2+(a^4+a^3+a^2+a)/4*x-x-a^3/2-a/2-1);
<algebraic extension over the Rationals of degree 10>
gap> x := Indeterminate(B,"x");
x
```

```
gap> Factor(x^5-5*x+12,B);
[ x+(!-a), x+(-a), x+(a+(1/4*a^4+1/4*a^3+1/4*a^2+1/4*a-1)),
  x+((1/2*a+1/2)*a+(1/2*a-1/2)),
  x+((-1/2*a-1/2)*a+(-1/4*a^4-1/4*a^3-1/4*a^2+1/4*a+3/2)) ]
gap> ViewFactors(last,B,["a","b"]);
[ x-a, x-b, !!1/4*a^4+!!1/4*a^3+!!1/4*a^2+x+!!1/4*a+b-!!1,
  !!1/2*a*b+x+!!1/2*a+!!1/2*b+(!!-1/2),
  (!!-1/4)*a^4+(!!-1/4)*a^3+(!!-1/4)*a^2+(!!-1/2)*a*b+x+!!1/4*a+
      (!!-1/2)*b+!!3/2 ]
```

Yet if we had let b be a root of the *other* quadratic, would we get the same splitting field? The answer is yes, since the splitting fields are uniquely determined up to isomorphism. In order to prove this by induction, we actually have to prove slightly more.

PROPOSITION 14.8

Let ϕ be an isomorphism from the field F to a field E. Let

$$f(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots + c_nx^n$$

be a polynomial in $F[x]$. Then

$$g(x) = \phi(c_0) + \phi(c_1)x + \phi(c_2)x^2 + \phi(c_3)cx^3 + \cdots + \phi(c_n)x^n$$

is a polynomial in $E[x]$. Suppose that K is a splitting field of $f(x)$ over F, and L is a splitting field of $g(x)$ over E. Then there is an isomorphism μ from K to L, such that $\mu(x) = \phi(x)$ for all x in F.

PROOF If $f(x)$ has degree 1, then the roots of $f(x)$ are in F, and the roots of $g(x)$ are in E. Thus, $K = E$, and $L = F$, and so the function $\mu(x) = \phi(x)$ satisfies the necessary conditions.

Let us use induction on the degree of the polynomial $f(x)$. That is, we will assume that the proposition is true for all polynomials of degree $(n-1)$. By lemma 13.3, the isomorphism ϕ extends to an isomorphism from $F[x]$ to $E[x]$ in such a way that $\phi(x) = x$. Thus, if $p(x)$ is an irreducible factor of the polynomial $f(x)$, then $\phi(p(x))$ is an irreducible factor of the polynomial $g(x) = \phi(f(x))$. Note that every root of $p(x)$ is also a root of $f(x)$, so that $p(x)$ factors completely in the field K. Likewise, $\phi(p(x))$ factors completely in the field L.

Let u be a root of $p(x)$ in K, and let v be a root of $\phi(p(x))$ in L. By proposition 14.7, there is an isomorphism θ mapping $F(u)$ to $E(v)$, such that $\theta(u) = v$, and $\theta(x) = \phi(x)$ for all x in F.

Since u is a root of $f(x)$, we can write $f(x) = (x-u) \cdot h(x)$, with $h(x)$ in $F(u)[x]$. Then

$$g(x) = \phi(f(x)) = \theta(f(x)) = \theta(x-u) \cdot \theta(h(x)) = (x-v) \cdot \theta(h(x)).$$

Since $h(x)$ has degree $(n-1)$, we can use the induction hypothesis. Obviously K is the splitting field of $h(x)$ over $F(u)$, and L is the splitting field of $\theta(h(x))$ over $E(v)$. Thus, by the induction hypothesis the proposition is true for the polynomial $h(x)$, so there is an isomorphism μ such that $\mu(x) = \theta(x)$ for all x in $F(u)$. Since $\theta(x) = \phi(x)$ for all x in F, we have found an isomorphism with the necessary properties. □

COROLLARY 14.3
If $f(x)$ is a polynomial in $F[x]$, then all splitting fields of $f(x)$ are isomorphic.

PROOF Simply let $F = E$, and let $\phi(x) = x$ for all x in F. Then by proposition 14.8, any two splitting fields of $f(x) = g(x)$ will be isomorphic. □

In section 13.3, we studied the properties of cyclotomic polynomials. It will be important later on to determine the splitting fields of these polynomials. For example, the ninth cyclotomic polynomial is given as $x^6 + x^3 + 1$. The splitting field found by GAP or *Mathematica* is only 6-dimensional—the splitting field is simply $\mathbb{Q}(a)$, where a is one root of the polynomial.

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals, x^6 + x^3 + 1);
<algebraic extension over the Rationals of degree 6>
gap> x := Indeterminate(A,"x");
x
gap> Factor(x^6 + x^3 + 1, A);
[ x+(-a), x+(-a^2), x+(-a^4), x+(-a^5), x+(a^4+a), x+(a^5+a^2) ]
```

We can quickly generalize this result to apply to all cyclotomic polynomials.

PROPOSITION 14.9
The splitting field of the n-th cyclotomic polynomial has dimension at most $\phi(n)$ over $\mathbb{Q}$, where $\phi(n)$ is Euler's totient function. In fact, the splitting field is given as $\mathbb{Q}(e_n)$, where e_n is a primitive n-th root of unity.

PROOF From the definition of the splitting field, the generator

$$e_n = e^{(2\pi i/n)} = \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right)$$

is a root of the n-th cyclotomic polynomial

$$\Phi_n(x) = (x - (e_n)^{(k_1)}) \cdot (x - (e_n)^{(k_2)}) \cdot (x - (e_n)^{(k_3)}) \cdot \cdots \cdot (x - (e_n)^{(k_i)}),$$

where $k_1, k_2, k_3, \ldots k_i$ are the integers from 1 to n that are coprime to n. Thus, the splitting field contains $\mathbb{Q}(e_n)$. Note that all powers of e_n are in this

field, and so the n-th cyclotomic polynomial factors completely in $\mathbb{Q}(e_n)$. To find the dimension of $\mathbb{Q}(e_n)$ over $\mathbb{Q}$, we first let $g(x) = \text{Irr}_{\mathbb{Q}}(e_n, x)$, and use corollary 14.1 to show that the dimension of $\mathbb{Q}(e_n)$ over $\mathbb{Q}$ is the degree of $g(x)$. But e_n is a root of $\Phi_n(x)$, which has dimension $\phi(n)$ and is in $\mathbb{Q}[x]$ by corollary 13.4. So the degree of $g(x)$ is at most $\phi(n)$. Therefore, the dimension of the splitting field of $\Phi_n(x)$ is at most $\phi(n)$. □

In fact, the n-th cyclotomic polynomial will always be irreducible, so in fact the splitting field of $\Phi_n(x)$ will in fact have dimension $\phi(n)$ over $\mathbb{Q}$. However, we never officially proved that these polynomials are all irreducible.

We now will show that splitting fields have special properties that most field extensions do not have. For example, we can define the splitting field of $x^3 - 2$ as follows:

InitDomain[0]
Define[a^3, 2]
Define[b^2, − a^2 − a b]

Note that $x^2 + 3$ factors in the splitting field, as does $x^6 + 108$. In fact, both polynomials factor completely in this field $\mathbb{Q}(a, b)$.

Factor[x^2 + 3, a, b]
Factor[x^6 + 108, a, b]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals,x^3-2);
<algebraic extension over the Rationals of degree 3>
gap> x := Indeterminate(A,"x");
x
gap> a := PrimitiveElement(A);
a
gap> B := FieldExtension(A,x^2+a*x+a^2);
<algebraic extension over the Rationals of degree 6>
gap> x := Indeterminate(B,"x");
x
gap> Factor(x^2+3,B);
[ x+((-a^2)*a-!1), x+(a^2*a+!1) ]
gap> ViewFactors(last,B,["a","b"]);
[ -a^2*b+x-!!1, a^2*b+x+!!1 ]
gap> Factor(x^6+108,B);
[ x+(a+2*a), x+(!2*a+a), x+(a+(-a)), x+(-a+(-2*a)),
  x+((!-2)*a+(-a)), x+(-a+a) ]
gap> ViewFactors(last,B,["a","b"]);
[ x+!!2*a+b, x+a+!!2*b, x-a+b, x+(!!-2)*a-b, x-a+(!!-2)*b,
  x+a-b ]
```

This last example suggests a startling fact: Whenever an irreducible polynomial in $\mathbb{Q}[x]$ has just one root in a splitting field, then the polynomial factors completely in the splitting field. This property characterizes splitting fields from other extensions of $\mathbb{Q}$.

LEMMA 14.5
Let K be the splitting field of a polynomial $f(x)$ in $F[x]$. Then if $p(x)$ is an irreducible polynomial in $F[x]$ for which there is one root in K, then $p(x)$ factors completely in K.

PROOF Let $u_1, u_2, u_3, \ldots, u_n$ be the roots of $f(x)$ in K. Then

$$K = F(u_1, u_2, u_3, \ldots, u_n).$$

Suppose that $p(x)$ has one root v in K. Consider $p(x)$ as a polynomial in K, and let L be the splitting field of $p(x)$ over K. Let w be any other root of $p(x)$ in L besides v. To show that $K = L$, we need to show that w is in K, which would show that all roots of $p(x)$ are in K.

By proposition 14.7, there is an isomorphism ϕ from $F(v)$ to $F(w)$ such that $\phi(v) = w$, and $\phi(x) = x$ for all x in F. (We let $f(x) = x$, the identity map, and let E and K both be the field F.) By lemma 13.3 we can extend ϕ to an isomorphism from $F(v)[x]$ to $F(w)[x]$, and $\phi(f(x)) = f(x)$.

We now want to consider the field $K(w)$. We have

$$K(w) = F(u_1, u_2, u_3, \ldots, u_n, w) = F(w, u_1, u_2, u_3, \ldots, u_n).$$

Thus, $K(w)$ is the splitting field of $f(x)$ over the field $F(w)$. Since v is in K,

$$K = K(v) = F(u_1, u_2, u_3, \ldots, u_n, v) = F(v, u_1, u_2, u_3, \ldots, u_n),$$

so K is the splitting field of $f(x)$ over the field $F(v)$.

Consequently proposition 14.8 shows us that the isomorphism ϕ from $F(v)$ to $F(w)$ extends to an isomorphism μ from K to $K(w)$, and $\mu(v) = w$. Also, $\mu(x) = x$ for all x in F. Thus, we can use corollary 14.3 to show that K and $K(w)$ have the same dimension over F. By proposition 14.2, the dimension of $K(w)$ over F equals the dimension of $K(w)$ over K times the dimension of K over F. Therefore, the dimension of $K(w)$ over K must be 1, so w is in K. Therefore, every root of $p(x)$ is in K, so $p(x)$ factors completely in K. ∎

The fact that the splitting field of $x^6 + 108$ is the same as the splitting field of $x^3 - 2$ reveals another curious property of splitting fields. Rather than having to make an "extension of an extension" to define the splitting field $\mathbb{Q}(a, b)$, we could have defined the same field using a single extension of the element $w = \sqrt[6]{-108}$.

DEFINITION 14.9 We say that a finite extension of a field K is called a *simple extension* if it can be expressed as $K(a)$ for some element a.

The splitting field of $x^3 - 2$, even though it was originally described as an extension of an extension, is in fact a simple extension of $\mathbb{Q}$ of dimension 6.

Let us show, using the splitting fields, that an extension of an extension will usually form a simple extension.

PROPOSITION 14.10

Let F be a field, and let K be a finite-dimensional extension of F. Suppose that $K = F(u, v)$ with u, v in K. Let L be the splitting field of the polynomial $g(x) = \mathrm{Irr}_F(v, x)$, and suppose that there are no multiple roots of $g(x)$ in the field L. Then there is an element w of K such that $K = F(w)$.

PROOF If F is a finite field, then K will also be a finite field, and we can simply let w be a generator of the multiplicative group K^*, using proposition 13.4. Thus, we will assume that F is an infinite field. Let $f(x) = \mathrm{Irr}_F(u, x)$ and $g(x) = \mathrm{Irr}_F(v, x)$. Let E be the splitting field of $g(x)$ over the field $F(u)$. Since $g(x)$ factors completely in L without double roots, $g(x)$ will also factor completely in E without double roots. Let $v = v_1, v_2, v_3, \ldots, v_k$ be the distinct roots of $g(x)$ in E.

Since u is in E, there is at least one root of $f(x)$ in the field E. Even though $f(x)$ may not factor completely in the field E we can let $u = u_1, u_2, u_3, \ldots, u_n$ be the roots of $f(x)$ over E.

Since F is an infinite field, we can pick some element y of F, such that

$$y \neq \frac{u_i - u}{v - v_j} \qquad \text{for all } 1 \leq i \leq n, \quad 1 < j \leq k.$$

Finally, we let $w = u + yv$. Let us show that $K = F(w)$. To show that v is in $F(w)$, let $h(x) = f(w - yx)$, and note that $h(v) = f(u + yv - yv) = f(u) = 0$ so v is a root of $h(x)$. If one of the other roots of $g(x)$ is a root of $h(x)$, then $w - yv_j = u + yv - yv_j = u_i$ for some j and i, which would give us

$$y = \frac{u_i - u}{v - v_j},$$

and we specifically chose y so that it would avoid these values. Thus, there is only one root in common between $g(x)$ and $h(x)$ in the field E.

Let $r(x) = \mathrm{Irr}_{F(w)}(v, x)$. Then $r(x)$ divides the polynomials $g(x)$ and $h(x)$, since both polynomials have v as a root. In fact, we have seen that $g(x)$ and $h(x)$ have no other roots in common, so $r(x)$ has only one root in the field E. But $g(x)$ splits completely in E, and has no multiple roots in E. Thus, $r(x)$ has degree 1, and in fact $r(x) = x - v$. This proves that v is in $F(w)$. To see that u is in $F(w)$, we note that $u = yv - w$. Thus, $F(u, v)$ is contained in $F(w)$ while $F(w)$ is obviously contained in $F(u, v)$. Therefore, $F(u, v) = F(w)$. □

COROLLARY 14.4

Let K be a finite-dimensional extension of F, with $K = F(u_1, u_2, u_3, \ldots u_n)$ and suppose that none of the polynomials $\mathrm{Irr}_F(u_i)$ have multiple roots in each

of their splitting fields. Then there exists an element w in K such that $K = F(w)$.

PROOF We will proceed by induction on n. If $n = 1$, we can let $w = u_1$, and there is nothing to prove. If $n = 2$ we can use proposition 14.10 to find w. Suppose that the corollary is true for the previous case, so that we found a u is K such that $F(u) = F(u_1, u_2, u_3, \ldots, u_{n-1})$. Let $v = u_n$, and since $g(x) = \mathrm{Irr}_R(u_{k+1})$ does not have a multiple root in its splitting field L, we can use proposition 14.10 to find a w in K such that $F(w) = F(u, v)$. But then $F(w) = F(u_1, u_2, u_3, \ldots, u_{n-1}, u_n)$. Thus, the corollary is true for all positive values of n. ∎

Mathematica and GAP have a function **SimpleExtension** that finds one of the many elements w for which the field $\mathbb{Q}(a, b, \ldots) = \mathbb{Q}(w)$. For example, the splitting field of $x^3 - 2$ is $\mathbb{Q}(a, b)$, which is defined above. We then can find an element w by the command

SimpleExtension[a,b]

which returns $a + 2b$. Thus, $\mathbb{Q}(a, b) = \mathbb{Q}(a + 2b)$, which is a simple extension. This element turns out to be a sixth root of -108. GAP's `SimpleExtension` does even more. It forms a new field extension, which is a simple extension, and then provides a way to map back and forth from the original field to the new field. For example, using the B defined above,

```
gap> L := SimpleExtension(B);
[ <algebraic extension over the Rationals of degree 6>,
  !2*a+a, [ 1/18*a^4, -1/36*a^4+1/2*a ] ]
gap> ViewFactors(last[2],B,["a","b"]);
a+!!2*b
```

This produces a new field, in which the new primitive element is $w = a + 2b$, and $a = w^4/18$ and $b = w/2 - w^4/36$. How does this command work? The key is in the proof of proposition 14.10. Within the proof, we found that $F(u, v) = F(u + yv)$, where y is any number such that

$$y \neq \frac{u_i - u}{v - v_j}$$

whenever u_i is a root of $\mathrm{Irr}_F(u, x)$, and v_j is a root of $\mathrm{Irr}_F(v, x)$.

Let us try another example. Consider $\mathbb{Q}(\sqrt[3]{2}, \sqrt{2})$. This is not a splitting field, but it is contained in the splitting field of $f(x) = (x^3 - 2)(x^2 - 2)$, which does not have multiple roots, so we can still apply proposition 14.10 to show that $\mathbb{Q}(\sqrt[3]{2}, \sqrt{2}) = \mathbb{Q}(w)$ for some element w. But what is that element?

Note that $\mathrm{Irr}_{\mathbb{Q}}(\sqrt[3]{2}) = x^3 - 2$, which has roots of $\sqrt[3]{2}$, $e_3\sqrt[3]{2}$, and $e_3^2\sqrt[3]{2}$. Likewise, $\mathrm{Irr}_{\mathbb{Q}}(\sqrt{2}) = x^2 - 2$, which has roots of $\pm\sqrt{2}$. Hence, we must pick a

rational value of y that is not equal to

$$\frac{e_3^i \sqrt[3]{2} - \sqrt[3]{2}}{\sqrt{2} \pm \sqrt{2}}.$$

That is, y cannot equal 0, $(e_3 - 1)\sqrt[3]{2}/(2\sqrt{2})$, or $(e_3^2 - 1)\sqrt[3]{2}/(2\sqrt{2})$. Any other rational value of y will do, so for convenience we can take $y = 1$. Then $w = u + yv = \sqrt[3]{2} + \sqrt{2}$.

We can also have *Mathematica* or GAP find an element for us.

InitDomain[0]
Define[a^3, 2]
Define[b^2, 2]
SimpleExtension[a,b]
$a + b$

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals,x^3-2);
<algebraic extension over the Rationals of degree 3>
gap> x := Indeterminate(A,"x");
x
gap> B := FieldExtension(A,x^2-2);
<algebraic extension over the Rationals of degree 6>
gap> SimpleExtension(B);
[ <algebraic extension over the Rationals of degree 6>, a+a,
  [ -12/155*a^5-9/310*a^4+16/31*a^3+78/155*a^2-76/155*a+182/155,
  12/155*a^5+9/310*a^4-16/31*a^3-78/155*a^2+231/155*a-182/155] ]
```

There is in fact an easier way to find a simple extension in this case. Merely note that $\sqrt[6]{2} \in \mathbb{Q}(\sqrt[3]{2}, \sqrt{2})$, since $\sqrt[6]{2} = \sqrt{2}/\sqrt[3]{2}$. Yet $\sqrt{2} = \sqrt[6]{2}^3$, and $\sqrt[3]{2} = \sqrt[6]{2}^2$. So $\mathbb{Q}(\sqrt[3]{2}, \sqrt{2}) = \mathbb{Q}(\sqrt[6]{2})$.

The fact that we can convert an extension of an extension to a simple extension will simplify many of the proofs involving splitting fields. In particular, it will allow us to explore the automorphisms of the splitting fields. In the next chapter we will discover that the automorphisms of the splitting fields determine much of the information about the roots of the polynomial, and whether they can be expressed in terms of square roots and cube roots. This beautiful correlation is referred to as Galois theory.

Problems for Chapter 14

Interactive Problems

14.1 Use *Mathematica* or GAP to find the coefficients of the vector $\langle 3, -2, 5\rangle$ in $\mathbb{R}^3$ using the basis $\{\langle 2, -1, 4\rangle, \langle 5, 2, 1\rangle, \langle 4, -3, 2\rangle\}$.

14.2 Use *Mathematica* or GAP to find the coefficients of the element $Z(27)^5$ in GF(27) over Z_3 using the basis $\{Z(3)^0, Z(27), Z(27)^2\}$. Note that the Conway polynomial of degree 3 over Z_3 is $x^3 + 2x + 1$.

14.3 Define the field $\mathbb{Q}(\sqrt{-3})$ in GAP or *Mathematica*, then find $1/(5 + \sqrt{-3})$. Note that in *Mathematica*, you must first use the **CheckField** command to show that $\{1, \sqrt{-3}\}$ is a basis.

14.4 Define the field $\mathbb{Q}(\sqrt{5})$ in GAP or *Mathematica*. Does the polynomial $x^2 + 4x - 1$ factor in this field?

For problems **14.5** through **14.8**: Define the splitting field of the polynomial in *Mathematica* or GAP. Determine the dimension of the splitting field over $\mathbb{Q}$.

14.5 $x^3 + x^2 - 4x + 1$
14.6 $x^5 + x^4 - 4x^3 - 3x^2 + 3x + 1$
14.7 $x^5 - 2$
14.8 $x^5 + 20x + 16$

Non-Interactive Problems

For problems **14.9** through **14.14**: Find a basis for the following fields over $\mathbb{Q}$.

14.9 $\mathbb{Q}(\sqrt{2})$
14.10 $\mathbb{Q}(\sqrt{5})$
14.11 $\mathbb{Q}(\sqrt{2}, \sqrt{3})$
14.12 $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$
14.13 $\mathbb{Q}(\sqrt[3]{2})$
14.14 $\mathbb{Q}(e_9)$

14.15 Find a basis for the field $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ over the field $\mathbb{Q}(\sqrt{2})$.

For problems **14.16** through **14.21**: Find the following irreducible polynomials $\mathrm{Irr}_{\mathbb{Q}}(y, x)$.

Hint: Set $x = y$, and work to eliminate the roots.

14.16 $\mathrm{Irr}_{\mathbb{Q}}(\sqrt{5}, x)$
14.17 $\mathrm{Irr}_{\mathbb{Q}}(\sqrt[3]{5}, x)$
14.18 $\mathrm{Irr}_{\mathbb{Q}}(\sqrt{2} + \sqrt{3}, x)$
14.19 $\mathrm{Irr}_{\mathbb{Q}}\left(\sqrt{\sqrt{2} - 1}, x\right)$
14.20 $\mathrm{Irr}_{\mathbb{Q}}\left(\sqrt[3]{\sqrt{5} - 1}, x\right)$
14.21 $\mathrm{Irr}_{\mathbb{Q}}\left(\sqrt{\sqrt{\sqrt{2} - 1} + 1}, x\right)$

For problems **14.22** through **14.25**: Find all of the roots of the polynomial.

14.22 $\mathrm{Irr}_{\mathbb{Q}}(\sqrt{2} + \sqrt{3}, x)$. (See problem **14.18**.)
14.23 $\mathrm{Irr}_{\mathbb{Q}}\left(\sqrt{\sqrt{2} - 1}, x\right)$. (See problem **14.19**.)
14.24 $\mathrm{Irr}_{\mathbb{Q}}\left(\sqrt[3]{\sqrt{5} - 1}, x\right)$. (See problem **14.20**.)
14.25 $\mathrm{Irr}_{\mathbb{Q}}\left(\sqrt{\sqrt{\sqrt{2} - 1} + 1}, x\right)$. (See problem **14.21**.)

For problems **14.26** through **14.29**: Find a single number w such that the following field can be written as $\mathbb{Q}(w)$.

14.26 $\mathbb{Q}(\sqrt{2}, \sqrt[5]{2})$
14.27 $\mathbb{Q}(\sqrt{2}, \sqrt{5})$
14.28 $\mathbb{Q}(\sqrt{2}, \sqrt{3}, \sqrt{5})$
14.29 $\mathbb{Q}(\sqrt[3]{2}, i)$
14.30 $\mathbb{Q}(\sqrt[3]{2}, e_3)$
14.31 $\mathbb{Q}(e_3, e_5)$

14.32 Show by direct computation that if a and b are two distinct roots of $x^3 - 2$, then $(a + 2b)^6 = -108$.

Hint: Use the fact that $b^2 = -ab - a^2$ to simplify as you go along.

14.33 Use either a calculator's *Solve* function or De Moivre's theorem (11.2) to find decimal approximations of the three roots of $x^3 - 2 = 0$. Verify that $a^2 + ab + b^2 = 0$ whenever a and b are two of the three roots.

14.34 The polynomial $x^3 + x - 1$ has one real root $a \approx 0.6823278038\ldots$. Show that the splitting field of this polynomial is 6-dimensional over $\mathbb{Q}$.

Hint: If $(x - a)$ is one factor, what is the other? Show that this other factor is irreducible in $\mathbb{R}$, and hence is irreducible in $\mathbb{Q}(a)$.

14.35 Find the splitting field of $x^4 + x^2 + 1 = (x^2 + x + 1)(x^2 - x + 1)$.

14.36 Let $F = Z_2(t)$ be the rational functions of t modulo 2. Let K be the splitting field of $x^2 - t$ (that is, $K = F(\sqrt{t})$). Show that K is isomorphic to F, even though K is an extension of F of order 2.

Hint: Let ϕ be a homomorphism that sends $\sqrt{t}$ to t.

14.37 Find the multiplicative inverse of $\sqrt[3]{4} - \sqrt[3]{2} - 3$ in $\mathbb{Q}(\sqrt[3]{2})$.

14.38 Let a be a root of the equation

$$x^5 + \sqrt{2}x^3 + \sqrt{3}x^2 + \sqrt{5}x + \sqrt{7}.$$

Show that $\mathbb{Q}(a)$ is a finite extension of $\mathbb{Q}$ with dimension at most 80.

14.39 Let K be a finite extension of a field F. If u and v are in K, prove that $F(u)(v) = F(v)(u)$.

14.40 Suppose $f(x)$ and $g(x)$ are two polynomials in $\mathbb{Q}[x]$. Suppose that the splitting field of $f(x)$ is of dimension n over $\mathbb{Q}$, and the splitting field of $g(x)$ is of dimension m over $\mathbb{Q}$. Prove that the splitting field of $f(x) \cdot g(x)$ has dimension no more than $n \cdot m$.

14.41 Let m and n be distinct integers. Show directly that $\mathbb{Q}(\sqrt{m}, \sqrt{n}) = \mathbb{Q}(\sqrt{m} + \sqrt{n})$.

Hint: $(\sqrt{m} + \sqrt{n})$, $(\sqrt{m} + \sqrt{n})^2$, and $(\sqrt{m} + \sqrt{n})^3$ are all in $\mathbb{Q}(\sqrt{m} + \sqrt{n})$. Find a way of obtaining $\sqrt{m}$ and $\sqrt{n}$ from these three expressions.

14.42 Prove that $\mathbb{Q}(\sqrt{2})$ is not isomorphic to $\mathbb{Q}(\sqrt{3})$.

14.43 Find all of the automorphisms of $\mathbb{Q}(\sqrt{2}, \sqrt{3})$.

This page intentionally left blank

Chapter 15

Galois Theory

15.1 The Galois Group of an Extension Field

In the last chapter, we explored the extensions of a field, and found that any finite extension could be entered into *Mathematica*® fairly easily. In particular, we explored the splitting fields of several polynomials. In this chapter, we will explore the automorphisms on the field extensions, and discover that the group of automorphisms contains much information about the polynomial. For example, it will tell us if the roots of the polynomial can be expressed in terms of square roots and cube roots.

DEFINITION 15.1 Let K be a finite extension of the field F. An *F-automorphism* of K is a ring automorphism ϕ on the field K that fixes every element of F. That is, $\phi(x) = x$ whenever x is in F.

Note that there is at least one F-automorphism of K, the identity automorphism. Since we have seen that the set of group automorphisms of a group forms another group, it is not surprising that the same thing happens for F-automorphisms of a field.

PROPOSITION 15.1

If K is a finite extension of a field F, then the set of all F-automorphisms of K forms a group under the operation of composition of functions.

PROOF By lemma 11.5, the set of all ring automorphisms of a ring forms a group. So we only need to show that the set of F-automorphisms of K is a subgroup of the group of all automorphisms. If ϕ_1 and ϕ_2 are two F-automorphisms of K, then $\phi_1(x) = \phi_2(x) = x$ for all x in F. Thus, $(\phi_1 \cdot \phi_2)(x) = \phi_2(\phi_1(x)) = x$ for all x in F. Thus, $\phi_1 \cdot \phi_2$ is an F-automorphism of K. Note also that $\phi_1^{-1}(x) = x$ for all x in F, so ϕ_1^{-1} is also an F-automorphism of K. Since the set of all F-automorphisms of K is closed under multiplications and inverses, this set is a group. ∎

DEFINITION 15.2 The set of all F-automorphisms of K is denoted $\mathrm{Gal}_F(K)$, and is called the *Galois Group of K over F.*

For example, the set of complex numbers $\mathbb{C}$, according to proposition 11.4, has two automorphisms that fix the real numbers: the identity automorphism, and the automorphism that sends each number to its complex conjugate. So there are exactly two elements of $\mathrm{Gal}_{\mathbb{R}}(\mathbb{C})$. In other words, $\mathrm{Gal}_{\mathbb{R}}(\mathbb{C})$ is isomorphic to Z_2.

We want to find a way to compute the Galois group of any finite extension of a field F. Since we can define finite extensions in terms of polynomials, it is natural to ask what must happen to the roots of a polynomial.

LEMMA 15.1
Let K be a finite extension of F, and let $f(x)$ be a polynomial in $F[x]$. If u is a root of $f(x)$, and ϕ is in $\mathrm{Gal}_F(K)$, then $\phi(u)$ is also a root of $f(x)$.

PROOF Let $f(x) = c_0 + c_1x + c_2x^2 + c_3x^3 + \cdots + c_nx^n$. Since u is a root of $f(x)$ we have that

$$c_0 + c_1u + c_2u^2 + c_3u^3 + \cdots + c_nu^n = 0.$$

Since ϕ is a ring homomorphism, we have that

$$\begin{aligned} 0 = \phi(0) &= \phi(c_0 + c_1u + c_2u^2 + c_3u^3 + \cdots + c_nu^n) \\ &= \phi(c_0) + \phi(c_1)\phi(u) + \phi(c_2)\phi(u^2) + \phi(c_3)\phi(u^3) + \cdots + \phi(c_n)\phi(u^n). \end{aligned}$$

Since $c_0, c_1, c_2, \ldots c_n$ are in F, we have

$$0 = c_0 + c_1\phi(u) + c_2\phi(u)^2 + c_3\phi(u)^3 + \cdots + c_n\phi(u)^n.$$

Therefore, $\phi(u)$ is also a root of $f(x)$. □

Let us use this lemma to find the Galois group of the splitting field of x^3-2. The splitting field is defined by letting $a^3 = 2$, and $b^2 = -a^2 - ab$.

InitDomain[0]
Define[a^3, 2]
Define[b^2, − a^2 − a b]
Factor[x^3 − 2, a, b]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals,x^3-2);
<algebraic extension over the Rationals of degree 3>
gap> a := PrimitiveElement(A);
a
gap> x := Indeterminate(A,"x");
x
```

```
gap> B := FieldExtension(A,x^2 + a*x + a^2);
<algebraic extension over the Rationals of degree 6>
gap> x := Indeterminate(B,"x");
x
gap> Factor(x^3-2,B);
[ x+(!-a), x+(-a), x+(a+a) ]
gap> ViewFactors(last,B,["a","b"]);
[ x-a, x-b, x+a+b ]
```

The three roots of $x^3 - 2$ are a, b, and $-a - b$. Thus, lemma 15.1 tells us that if $F(x)$ is an automorphism on $\mathbb{Q}(a, b)$, then $F(a)$ is either a, b, or $-a - b$, while $F(b)$ is either a, b, or $-a - b$. Let us try to find an automorphism such that $F(a) = b$ and $F(b) = a$.

Homomorph[F]
Define[F[a], b]
Define[F[b], a]
CheckHomo[F, {a, b}]

```
gap> b := PrimitiveElement(B);
a
gap> a := a*One(b);
!a
gap> F := AlgebraHomomorphismByImagesNC(B,B,[a,b],[b,a]);
[ !a, a ] -> [ a, !a ]
gap> CheckHomo(F,[a,b]);
true
```

We have successfully defined one automorphism of the Galois group. (Any nonzero homomorphism on a field must be an automorphism in light of proposition 10.5, and the fact that the kernel is always an ideal.) We can similarly define an automorphism $G(x)$ on $\mathbb{Q}(a, b)$ such that $G(a) = b$, and $G(b) = -a - b$.

```
gap> G := AlgebraHomomorphismByImagesNC(B,B,[a,b],[b,-a-b]);
[ !a, a ] -> [ a, -a+(-a) ]
gap> CheckHomo(G,[a,b]);
true
```

With these two automorphisms we can actually produce three more: $G(G(x))$, $F(G(x))$, and $G(F(x))$. *Mathematica* or GAP can show us that all five of these automorphisms are different, and if we include the identity automorphism, we have found six automorphisms on $\mathbb{Q}(a, b)$. Note that the Galois group is not abelian, since $F(G(x)) \neq G(F(x))$.

```
gap> Im(F,Im(G,a));
!a
gap> Im(F,Im(G,b));
-a+(-a)
gap> Im(G,Im(F,a));
-a+(-a)
gap> Im(G,Im(F,b));
a
```

This introduces a new GAP command `Im` that finds the image of a homomorphism at a particular element.

It seems as though we must have found all of the automorphisms at this point, but this still needs to be proved. We begin by showing that there will always be an automorphism that moves one root of an irreducible polynomial to another.

PROPOSITION 15.2

Let K be the splitting field of some polynomial $f(x)$ over F, and let u and v be two elements of K. Then there exists an F-automorphism ϕ such that $\phi(u) = v$ if, and only if, $\text{Irr}_F(u, x) = \text{Irr}_F(v, x)$.

PROOF If there is some ϕ such that $\phi(u) = v$, we can let $g(x) = \text{Irr}_F(u, x)$ and $h(x) = \text{Irr}_F(v, x)$. Then u is a root of $g(x)$, and v is a root of $h(x)$. By lemma 15.1, u is a root of $h(x)$ and v is a root of $g(x)$, since $v = \phi^{-1}(u)$. So $g(x)$ is a multiple of $h(x)$, and vice versa. Since both have a leading coefficient of 1, we have that $g(x) = h(x)$.

Now suppose that $\text{Irr}_F(u, x) = \text{Irr}_F(v, x)$. Then by corollary 14.2 there is an isomorphism ϕ from $F(u)$ to $F(v)$ such that $\phi(u) = v$, and $\phi(x) = x$ for all x in F. Since K is a splitting field of $f(x)$ over F, it is a splitting field of $f(x)$ over both $F(u)$ and $F(v)$. Therefore ϕ extends to an F-automorphism of K (which we will also denote ϕ) by proposition 14.8. Therefore, ϕ is in $\text{Gal}_F(K)$, and $\phi(u) = v$. □

The next lemma will be important in determining the subgroups of the Galois group.

LEMMA 15.2

Let K be a finite extension of F, and let ϕ be an F-automorphism of K. Then the set of all elements x such that $\phi(x) = x$ forms a subfield of K containing F.

PROOF Let E be the set of all elements x such that $\phi(x) = x$. Since ϕ is an F-automorphism, by definition E must contain the elements of F. If x and y are in E, note that

$$\phi(x + y) = \phi(x) + \phi(y) = x + y,$$

$$\phi(x \cdot y) = \phi(x) \cdot \phi(y) = x \cdot y,$$

$$\phi(-x) = -\phi(x) = -x,$$

$$\phi(x^{-1}) = \phi(x)^{-1} = x^{-1}, \qquad \text{if } x \neq 0.$$

Thus, $x + y$, $x \cdot y$, and $-x$ are in E whenever x and y are, and x^{-1} is in E whenever $x \neq 0$ is in E. Thus, E is a subfield of K. ∎

Next we want to work on finding an upper bound on the number of elements in $\mathrm{Gal}_F(K)$.

PROPOSITION 15.3

Let $K = F(u_1, u_2, u_3, \ldots, u_n)$ be a finite extension field of F. If ϕ_1 and ϕ_2 are two F-automorphisms in $\mathrm{Gal}_F(K)$, and

$$\phi_1(u_1) = \phi_2(u_1), \qquad \phi_1(u_2) = \phi_2(u_2), \qquad \ldots \quad \phi_1(u_n) = \phi_2(u_n),$$

then $\phi_1(x) = \phi_2(x)$ for all x in K. In other words, an F-automorphism in $\mathrm{Gal}_F(K)$ is completely determined by its action on $u_1, u_2, u_3, \ldots, u_n$.

PROOF Consider the F-automorphism $\phi_2^{-1}(\phi_1(x))$. It is clear that this automorphism fixes $u_1, u_2, u_3, \ldots u_n$, as well as the elements of F. By lemma 15.2, the set E of all elements x such that $\phi_2^{-1}(\phi_1(x)) = x$ forms a subfield of K. But K is by lemma 14.3 the smallest field containing u_1, u_2,$u_3, \ldots, u_n$, and F. Thus, $K = E$, and so $\phi_1(x) = \phi_2(x)$ for all x in K. ∎

We can now apply this proposition to the field $\mathbb{Q}(a, b)$. Any $\mathbb{Q}$-automorphism is determined by where it sends the elements a and b. By lemma 15.1, these elements can only be sent to a, b, or $-a - b$. Yet an automorphism cannot send two elements to the same element. Thus, there are at most six $\mathbb{Q}$-automorphisms on the field $\mathbb{Q}(a, b)$. Yet we have found precisely six $\mathbb{Q}$-automorphisms of $\mathbb{Q}(a, b)$. Thus, we have found all of the $\mathbb{Q}$-automorphisms, and the Galois group of $\mathbb{Q}(a, b)$ contains exactly six elements. Furthermore, we observed that $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(a, b))$ was non-commutative, so we find that $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(a, b))$ must be isomorphic to S_3.

We can find an upper bound for the number of F-automorphisms in any splitting field using a similar argument.

COROLLARY 15.1

If K is the splitting field of a polynomial $f(x)$ of degree n in $F[x]$, then $\mathrm{Gal}_F(K)$ is isomorphic to a subgroup of S_n.

PROOF Since $f(x)$ has degree n in $F[x]$, there are at most n roots of $f(x)$ in K. Call these roots $u_1, u_2, \ldots, u_m$. Since K is the splitting field of $f(x)$ over F, we can write $K = F(u_1, u_2, u_3, \ldots, u_m)$. If ϕ is in $\mathrm{Gal}_F(K)$, then $\phi(u_1), \phi(u_2), \phi(u_3), \ldots, \phi(u_m)$ will be distinct roots of $f(x)$ by lemma 15.1. Hence, ϕ will act as a permutation on the roots of $f(x)$. By proposition 15.3, ϕ is completely determined by this permutation on the roots of $f(x)$. Thus,

$\mathrm{Gal}_F(K)$ is isomorphic to a subgroup of S_m, and since m is not larger than n, $\mathrm{Gal}_F(K)$ is isomorphic to a subgroup of S_n. □

We immediately see from this corollary that the Galois group of a finite extension must be a finite group.

Let us look at one more example of a Galois group of a field. Consider the field $\mathbb{Q}(\sqrt[3]{2})$, which is a subfield of the field $\mathbb{Q}(a, b)$. Note that in this subfield all of the elements are *real*. Thus, in this field $\mathbb{Q}(\sqrt[3]{2})$ there is only one root to the polynomial $x^3 - 2$. Hence, if $\phi(x)$ is a $\mathbb{Q}$-automorphism of $\mathbb{Q}(\sqrt[3]{2})$, then $\phi(\sqrt[3]{2})$ must be $\sqrt[3]{2}$. By proposition 15.3, the $\mathbb{Q}$-automorphism is completely determined by where ϕ sends $\sqrt[3]{2}$. Thus, $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(\sqrt[3]{2}))$ is merely the trivial group.

In order to find the Galois group of a field, it is very helpful to know ahead of time the exact size of the Galois group. The next proposition allows us to compute the size of the Galois group for an important class of field extensions.

PROPOSITION 15.4

Suppose K is the splitting field of a polynomial $f(x)$ in $F[x]$, and that K can be expressed as a simple extension $K = F(w)$. If $\mathrm{Irr}_F(w, x)$ has no double roots in K, then the number of F-automorphisms in $\mathrm{Gal}_F(K)$ is precisely the dimension of K over F.

PROOF Let d be the dimension of K over F. Then if $g(x) = \mathrm{Irr}_F(w, x)$, then $g(x)$ has degree d. Since K is a splitting field and contains one root of $g(x)$, by lemma 14.5 $g(x)$ splits completely in K. Since there are no double roots of $g(x)$ in K, then there are d roots $w = w_1, w_2, w_3 \cdots w_d$. Since $g(x)$ is irreducible, $\mathrm{Irr}_F(w_i, x) = \mathrm{Irr}_F(w, x)$ so proposition 15.2 states that there is an F-automorphism that sends w to w_i for $1 \leq i \leq d$. Hence, there are at least d F-automorphisms. But by proposition 15.3, the F-automorphism of $F(w)$ is determined by where it sends w, which must be one of the d roots. So $|\mathrm{Gal}_F(K)| = d$. □

We are ready to try a more complicated example. Suppose we want to find the Galois group for the splitting field of the polynomial $x^4 - 2x^3 + x^2 + 1$. First we verify that this polynomial is irreducible.

Factor[x^4 − 2 x^3 + x^2 + 1]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Factor(x^4-2*x^3 + x^2 + 1,Rationals);
[ x^4-2*x^3+x^2+1 ]
```

Mathematica and GAP show this polynomial is irreducible over $\mathbb{Q}$. Let us define a to be one root of this polynomial, and see how this polynomial factors over $\mathbb{Q}(a)$.

InitDomain[0]
Define[a^4, 2 a^3 − a^2 − 1]
Factor[x^4 − 2 x^3 + x^2 + 1, a]
$(-a+x)(-1+a+x)(-a+a^2-x+x^2)$

Here is how we do this in GAP:

```
gap> A := FieldExtension(Rationals,x^4-2*x^3+x^2+1);
<algebraic extension over the Rationals of degree 4>
gap> x := Indeterminate(A,"x");
x
gap> Factor(x^4-2*x^3+x^2+1,A);
[ x+(-a), x+(a-1), x^2-x+(a^2-a) ]
```

This tells us that if a is a root, then $1-a$ is another root. However, it didn't factor completely, so we have to define b to be a root of the irreducible quadratic.

Define[b^2, b + a − a^2]
Factor[x^4 − 2 x^3 + x^2 + 1, a, b]
$(-a+x)(-1+a+x)(-b+x)(-1+b+x)$

```
gap> a := PrimitiveElement(A);
a
gap> B := FieldExtension(A,x^2-x+a^2-a);
<algebraic extension over the Rationals of degree 8>
gap> x := Indeterminate(B,"x");
x
gap> Factor(x^4-2*x^3+x^2+1,B);
[ x+(!-a), x+(-a), x+(a-!1), x+(!a-1) ]
gap> ViewFactors(last,B,["a","b"]);
[ x-a, x-b, x+b-!!1, x+a-!!1 ]
```

So the four roots are a, $1-a$, b, and $1-b$. Any $\mathbb{Q}$-automorphism will map each of these roots to another root, and so the Galois group will be a subgroup of S_4. But which permutations will give rise to a $\mathbb{Q}$-automorphism? A little trial and error will help.

Proposition 15.2 says that there will be some $\mathbb{Q}$-automorphism that sends any one of these four roots to any other of the four roots. So there is a $\mathbb{Q}$-automorphism that sends a to $1-a$. But where would it send the other three roots? Note that if $f(a) = 1-a$, then $f(1-a) = f(1) - f(a) = a$. So we only have to determine if $f(b)$ is b or $1-b$. *Mathematica* or GAP can show that both of these work, and *Mathematica* can draw a picture of how these two $\mathbb{Q}$-automorphisms act on the four roots of the polynomial.

```
gap> b := PrimitiveElement(B);
a
gap> a := a*One(b);
!a
gap> e := One(b);
!!1
gap> f := AlgebraHomomorphismByImagesNC(B,B,[a,b],[e-a,b]);
[ !a, a ] -> [ !-a+1, a ]
```

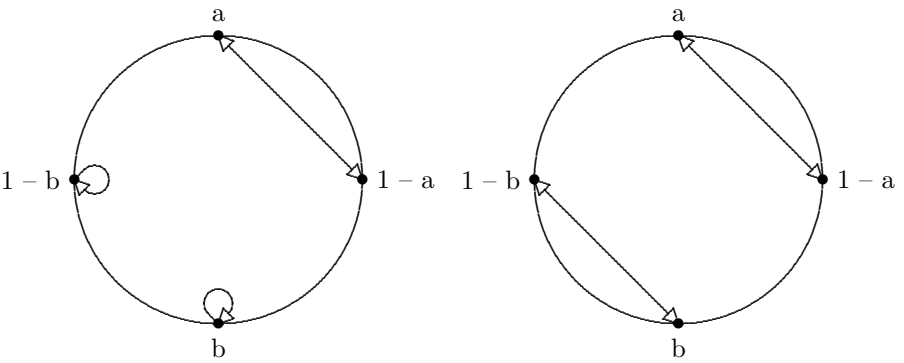


FIGURE 15.1: Two automorphisms of $\mathbb{Q}(a, b)$

```
gap> CheckHomo(f,[a,b]);
true
gap> g := AlgebraHomomorphismByImagesNC(B,B,[a,b],[e-a,e-b]);
[ !a, a ] -> [ !-a+1, -a+!1 ]
gap> CheckHomo(g,[a,b]);
true
```

Note that we had to define e to be the identity element of B, because `1-b` would cause a problem in GAP, since 1 is not an element of B.

If we number the four roots

$$1)\ a \qquad 2)\ 1-a \qquad 3)\ b \qquad 4)\ 1-b$$

we can view these two $\mathbb{Q}$-automorphisms as $P[2, 1]$ and $P[2, 1, 4, 3]$. The circle graphs of these two automorphisms are depicted in figure 15.1. But proposition 15.4 indicates that we must have eight $\mathbb{Q}$-automorphisms, so let us try mapping a to b. Then $1 - a$ would have to map to $1 - b$, but b could map to either a or $1 - a$. *Mathematica* shows that mapping b to a yields another $\mathbb{Q}$-automorphism, which would correspond to the permutation $P[3, 4, 1, 2]$. If we find the subgroup generated by these three $\mathbb{Q}$-automorphisms

M = Group[{P[2, 1], P[2, 1, 4, 3], P[3, 4, 1, 2]}]

```
gap> M := Group( (1,2),(1,2)(3,4), (1,3)(2,4) );
Group([ (1,2), (1,2)(3,4), (1,3)(2,4) ])
gap> Size(M);
8
```

we see that we have at least eight $\mathbb{Q}$-automorphisms. Since this is the number predicted by proposition 15.4, we are done. Hence, we found the Galois group as a subgroup of S_4 of order 8. The multiplication table

MultTable[M]

```
gap> NumberElements := true;
true
gap> MultTable(M);
*          |1  2  3  4  5  6  7  8
-----------+------------------------
()         |1  2  3  4  5  6  7  8
(1,2)      |2  1  4  3  6  5  8  7
(1,3)(2,4)|3  7  1  5  4  8  2  6
(1,4,2,3)  |4  8  2  6  3  7  1  5
(3,4)      |5  6  7  8  1  2  3  4
(1,2)(3,4)|6  5  8  7  2  1  4  3
(1,3,2,4)  |7  3  5  1  8  4  6  2
(1,4)(2,3)|8  4  6  2  7  3  5  1
```

shows that this group is non-abelian, and has five elements of order 2. Thus, the Galois group is isomorphic to D_4.

This example shows the usefulness of proposition 15.4 in finding the Galois group. In fact, sometimes the Galois group can be determined using only corollary 15.1 and proposition 15.4.

One of the tools we will use for finding the $\mathbb{Q}$-automorphisms is the close connection between the subgroups of the Galois group, and the subfields of the field extension. We begin by showing a way to produce subfields of a field extension using the subgroups of the Galois group.

PROPOSITION 15.5

Let K be a finite extension of F, and let H be a subgroup of $\mathrm{Gal}_F(K)$. *Let*

$$\mathrm{fix}(H) = \{k \in K \quad | \quad \phi(k) = k \qquad \text{for all } \phi \in H\}.$$

Then $\mathrm{fix}(H)$ *is a subfield of K containing the field F.*

PROOF For each ϕ in H, let E_ϕ be the set of elements that are fixed by ϕ. By lemma 15.2, E_ϕ is a subfield of K containing F. By taking the intersection of all E_ϕ with ϕ in H, we obtain a subfield of K containing F. □

DEFINITION 15.3 The field $\mathrm{fix}(H)$ is called the *fixed field* of the subgroup H.

Let us go back to the example of the Galois group of $\mathbb{Q}(a, b)$, where a and b were two roots of $x^3 - 2$.

The Galois group can be described as

$$\{I(x), F(x), G(x), G(G(x)), F(G(x)), G(F(x))\},$$

where $I(x)$ represents the identity automorphism that sends every element to itself. The subgroups of $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(a, b))$ are as follows:

$$H_1 = \{I(x)\}, \qquad H_2 = \{I(x), F(x)\}, \qquad H_3 = \{I(x), F(G(x))\},$$

$$H_4 = \{I(x), G(F(x))\}, \qquad H_5 = \{I(x), G(x), G(G(x))\},$$
$$H_6 = \{I(x), F(x), G(x), G(G(x)), F(G(x)), G(F(x))\}.$$

Let us find the six fixed fields of $\mathbb{Q}(a, b)$. The field $\text{fix}(H_1)$ is the set of elements fixed by the identity mapping, which is of course all of $\mathbb{Q}(a, b)$. The field $\text{fix}(H_2)$ contains the elements fixed by the mapping $F(x)$, which maps a to b, and b to a. Notice that the third root, $-a-b$, is fixed be the automorphism F. Thus, $\text{fix}(H_2) = \mathbb{Q}(-a-b)$. By a similar argument, we see that $\text{fix}(H_3) = \mathbb{Q}(a)$, and $\text{fix}(H_4) = \mathbb{Q}(b)$. The field $\text{fix}(H_5)$ is a little bit trickier, since $G(x)$ moves a, b, and $-a-b$. With a little bit of experimenting, we notice that

$$\begin{aligned} G(a^2b) &= b^2(-a-b) = (-a^2-ab)(-a-b) = a^3 + a^2b + a^2b + ab^2 \\ &= 2 + 2a^2b + a(-a^2-ab) = a^2b. \end{aligned}$$

If we substitute two of the roots of $x^3 - 2$ for a and b, that is, let $a = \sqrt[3]{2}$ and $b = e_3\sqrt[3]{2}$, we find that a^2b is $2e_3 = -1 + \sqrt{-3}$. This agrees with our previous observation that $\sqrt{-3}$ is in the field $\mathbb{Q}(a, b)$. Since -1 is already rational, we can write the fixed field $\text{fix}(H_5)$ as $\mathbb{Q}(\sqrt{-3})$.

Finally, the only elements of $\mathbb{Q}(a, b)$ that are fixed by *all* $\mathbb{Q}$-automorphisms are the elements of $\mathbb{Q}$. Hence $\text{fix}(H_6) = \mathbb{Q}$. Notice that we have found six different subfields of $\mathbb{Q}(a, b)$ by using the six subgroups of the Galois group. We will discover in the next section that this is *all* of the subfields of $\mathbb{Q}(a, b)$. Thus, we have found a convenient way of finding all of the subfields of a given field.

Here is another example, although a bit easier. Consider the field $\mathbb{Q}(\sqrt[3]{2})$. Since the only $\mathbb{Q}$-automorphism is the identity automorphism, which fixes the whole group, the only fixed field of $\mathbb{Q}(\sqrt[3]{2})$ is $\mathbb{Q}(\sqrt[3]{2})$, even though there is the obvious subfield $\mathbb{Q}$ within this field. We were hoping to be able to find *all* subfields of a field by looking at the fixed fields, but in this example we failed. We will understand why the field $\mathbb{Q}(\sqrt[3]{2})$ is not as well behaved as $\mathbb{Q}(a, b)$ in the next section.

15.2 The Galois Group of a Polynomial in $\mathbb{Q}$

To demonstrate Galois groups, let us concentrate on polynomials with rational coefficients. By working with rational numbers, we will avoid the problem of a splitting field having multiple roots. (In fields of finite characteristic, this can cause a problem.) This situation will never happen if we work in the field of rational numbers.

One advantage of working with a familiar field is that we can borrow a tool from calculus, namely the derivative. It isn't often that we will use a calculus result in algebra, but in this case it greatly simplifies the proof.

LEMMA 15.3
If $f(x)$ is an irreducible polynomial on $\mathbb{Q}[x]$, then $f(x)$ does not have multiple roots in the splitting field of $f(x)$.

PROOF Since we are working in $\mathbb{Q}[x]$, we can use the familiar tools of calculus. Suppose that K is the splitting field of $f(x)$, and u is a multiple root of $f(x)$ in K. Then

$$f(x) = (x-u)^2 \cdot g(x).$$

Since we are working in a field extension of $\mathbb{Q}$, we can take the derivative of both sides to get

$$f'(x) = 2(x-u) \cdot g(x) + (x-u)^2 g'(x).$$

Thus, u is a root of $f'(x)$, which has lower degree than $f(x)$. Note that $f'(x)$ is not 0, since it has degree of at least one.

Since $f'(x)$ is also in $\mathbb{Q}[x]$, we see that $\mathrm{Irr}_{\mathbb{Q}}[u,x]$ has degree less than the degree of $f(x)$, and so $\mathrm{Irr}_{\mathbb{Q}}[u,x]$ is a divisor of $f(x)$. But this contradicts the fact that $f(x)$ is irreducible. Therefore, $f(x)$ cannot have multiple roots in its splitting field. □

Because of this lemma, we know from proposition 14.10 that any splitting field can be expressed as a simple extension $\mathbb{Q}(w)$, and also we will be able to use proposition 15.4 to predict the size of the Galois group of the splitting field. We can relate the Galois group of the splitting field directly to the polynomial.

DEFINITION 15.4 Let $f(x)$ be a polynomial in $\mathbb{Q}$. The *Galois group* of $f(x)$ is the Galois group of the splitting field of $f(x)$ over $\mathbb{Q}$.

We have already seen some examples of Galois groups of splitting fields. The splitting field of $x^3 - 2$ was isomorphic to S_3. We also computed the Galois group of the splitting field of $x^4 - 2x^3 + x^2 + 1$, and found that the Galois group is isomorphic to D_4. Let us compute the Galois groups of some other polynomials.

Consider the polynomial $x^3 + x^2 - 2x - 1$. This polynomial is irreducible, as *Mathematica* or GAP can verify:

Factor[x^3 + x^2 − 2 x − 1]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> Factor(x^3+x^2-2*x-1,Rationals);
[ x^3+x^2-2*x-1 ]
```

Thus, we can let a denote one of the roots, and try to factor this in $\mathbb{Q}(a)$.

InitDomain[0]
Define[a^3, – a^2 + 2 a + 1]
Factor[x^3 + x^2 – 2 x – 1, a]

```
gap> A := FieldExtension(Rationals,x^3+x^2-2*x-1);
<algebraic extension over the Rationals of degree 3>
gap> x := Indeterminate(A,"x");
x
gap> Factor(x^3+x^2-2*x-1,A);
[ x+(-a), x+(-a^2+2), x+(a^2+a-1) ]
```

Since this factors completely, we see that the splitting field of x^3+x^2-2x-1 is $\mathbb{Q}(a)$. This is a 3-dimensional extension of $\mathbb{Q}$, so by proposition 15.4, the Galois group has three elements. Thus, the Galois group is isomorphic to Z_3.

Consider the polynomial $x^5-5x+12$. In the last chapter, we were able to find a splitting field by making two extensions, one of dimension 5, and one of dimension 2.

InitDomain[0]
Define[a^5, 5 a – 12]
Define[b^2, – 2 + 5 a/4 + a^2/4 + a^3/4 + a^4/4 – b – 3 a b/4 + a^2 b/4 + a^3 b/4 + a^4 b/4]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals,x^5-5*x+12);
<algebraic extension over the Rationals of degree 5>
gap> x := Indeterminate(A,"x");
x
gap> a := PrimitiveElement(A);
a
gap> e := One(A);
!1
gap> B := FieldExtension(A,(4*x^2+8*e-5*a-a^2-a^3-a^4+4*x+3*a*x
> -a^2*x-a^3*x-a^4*x)/(4*e));
<algebraic extension over the Rationals of degree 10>
```

If we define

c = Expand[(a^4 +a^3 + a^2–3 a–4 b–4) / 4]
d = Expand[
(a–4–a^2+a^3–a^4–4 b–a b+a^2 b–a^3 b +a^4 b)/8]
e = Expand[
(12–3a–a^2–3a^3–a^4+4b+a b–a^2 b+a^3 b–a^4 b)/8]

```
gap> x := Indeterminate(B,"x");
x
gap> b := PrimitiveElement(B);
a
gap> a := a*One(b);
!a
gap> e := One(b);
!!1
gap> c := (a^4 + a^3 + a^2 - 3*a - 4*b - 4*e)/(4*e);
```

```
-a+(1/4*a^4+1/4*a^3+1/4*a^2-3/4*a-1)
gap> d := (a-4*e-a^2+a^3-a^4-4*b-a*b+a^2*b-a^3*b+a^4*b)/(8*e);
(1/8*a^4-1/8*a^3+1/8*a^2-1/8*a-1/2)*a+
  (-1/8*a^4+1/8*a^3-1/8*a^2+1/8*a-1/2)
gap>e:=(12*e-3*a-a^2-3*a^3-a^4+4*b+a*b-a^2*b+a^3*b-a^4*b)/(8*e);
(-1/8*a^4+1/8*a^3-1/8*a^2+1/8*a+1/2)*a+
  (-1/8*a^4-3/8*a^3-1/8*a^2-3/8*a+3/2)
```

we see that the product

(x–a).(x–b).(x–c).(x–d).(x–e)

```
gap> (x-a)*(x-b)*(x-c)*(x-d)*(x-e);
x^5+(!!-5)*x+!!12
```

simplifies to $x^5 - 5x + 12$. Thus, the five roots are a, b, c, d, and e. (Note that we reused e, so this is no longer the identity element.) Any $\mathbb{Q}$-automorphism on the splitting field must send a and b to one of these five roots. Let us try to define a homomorphism f that sends $f(a) = b$, and $f(b) = a$.

Homomorph[F]
Define[F[a], b]
Define[F[b], a]
CheckHomo[F, {a, b}]

```
gap> f := AlgebraHomomorphismByImagesNC(B,B,[a,b],[b,a]);
[ !a, a ] -> [ a, !a ]
gap> CheckHomo(f,[a,b]);
true
```

Not only does *Mathematica* verify that this is a homomorphism, but it can also draw a circle graph describing how this homomorphism acts on the five roots. The left side of figure 15.2 is produced by the command

CircleGraph[{a, b, c, d, e}, F]

GAP is not able to form circle graphs, but the command `ChartHomo` will show where the five roots are mapped to. This command uses two lists: the first gives the roots, and the second gives the corresponding names for these roots.

```
gap> ChartHomo(f,[a,b,c,d,e],["a","b","c","d","e"]);
[ a -> b, b -> a, c -> d, d -> c, e -> e ]
```

Not every possible way of mapping a and b to the roots a, b, c, d, and e will produce a homomorphism. However, there is a homomorphism that maps $f(a) = a$ and $f(b) = c$. The commands

Homomorph[G]
Define[G[a], a]
Define[G[b], c]
CheckHomo[G, {a, b}]
CircleGraph[{a, b, c, d, e}, G]

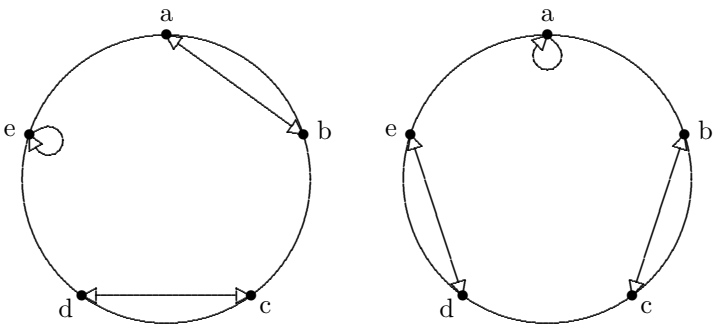


FIGURE 15.2: Two automorphisms for $x^5 - 5x + 12$

produce the right side of figure 15.2. In GAP, we must be content with just knowing where a, b, c, d, and e are mapped to.

```
gap> g := AlgebraHomomorphismByImagesNC(B,B,[a,b],[a,c]);
[ !a, a ] -> [ !a, -a+(1/4*a^4+1/4*a^3+1/4*a^2-3/4*a-1) ]
gap> CheckHomo(g,[a,b]);
true
gap> ChartHomo(g,[a,b,c,d,e],["a","b","c","d","e"]);
[ a -> a, b -> c, c -> b, d -> e, e -> d ]
```

Once we have found two $\mathbb{Q}$-automorphisms, we can find more by considering the group generated by these two elements. By corollary 15.1, the Galois group is a subgroup of S_5. We already have a natural ordering of the five roots, so the first permutation can be written **P[2,1,4,3]**, or $(1\,2)(3\,4)$, while the above permutation can be described as **P[1,3,2,5,4]**, or $(2\,3)(4\,5)$. Since the Galois group is a subgroup of S_5, we can ask *Mathematica* or GAP to find the subgroup generated by these two permutations.

G = Group[{P[2,1,4,3], P[1,3,2,5,4]}]

```
gap> G := Group((1,2)(3,4),(2,3)(4,5));
Group([ (1,2)(3,4), (2,3)(4,5) ])
gap> NumberElements := true;
true
gap> MultTable(G);

*           |1  2  3  4  5  6  7  8  9  10
------------+---------------------------
()          |1  2  3  4  5  6  7  8  9  10
(1,2,4,5,3)|2  3  4  5  1  10 6  7  8  9
(1,4,3,2,5)|3  4  5  1  2  9  10 6  7  8
(1,5,2,3,4)|4  5  1  2  3  8  9  10 6  7
(1,3,5,4,2)|5  1  2  3  4  7  8  9  10 6
(2,3)(4,5) |6  7  8  9  10 1  2  3  4  5
(1,2)(3,4) |7  8  9  10 6  5  1  2  3  4
(1,4)(3,5) |8  9  10 6  7  4  5  1  2  3
(1,5)(2,4) |9  10 6  7  8  3  4  5  1  2
(1,3)(2,5) |10 6  7  8  9  2  3  4  5  1
```

This produces exactly 10 permutations. Proposition 15.4 states that the size of the Galois group is equal to the dimension of the splitting field. Since the splitting field is a 2-dimensional extension of a 5-dimensional extension, the Galois group contains exactly 10 elements. Thus, we have found all of the $\mathbb{Q}$-automorphisms of the splitting field. The multiplication table of the Galois group reveals that the group is non-abelian. Since there is only one non-abelian group of order 10, the Galois group of $x^5 - 5x + 12$ is isomorphic to D_5.

Here is another example that illustrates the variety of groups that can be produced by a Galois group of a polynomial. Consider the eighth degree polynomial $x^8 - 24x^6 + 144x^4 - 288x^2 + 144$. This is an irreducible polynomial, as *Mathematica* or GAP can quickly verify. Thus, we can define a to be one root of this equation. GAP or *Mathematica* can then factor the polynomial in the field $\mathbb{Q}(a)$.

InitDomain[0]
Define[a^8, 24 a^6 – 144 a^4 + 288 a^2 – 144]
Factor[x^8 – 24 x^6 + 144 x^4 – 288 x^2 + 144, a]

```
gap> x := Indeterminate(Rationals,"x");
x
gap> A := FieldExtension(Rationals,
  x^8 - 24*x^6 + 144*x^4 - 288*x^2 + 144);
<algebraic extension over the Rationals of degree 8>
gap> x := Indeterminate(A,"x");
x
gap> Factor(x^8-24*x^6+144*x^4-288*x^2+144,A);
[ x+a, x+(1/24*a^7-5/6*a^5+5/2*a^3+a),
  x+(-1/12*a^5+3/2*a^3-3*a),x+(1/12*a^7-11/6*a^5+17/2*a^3-10*a),
  x+(-a), x+(-1/24*a^7+5/6*a^5-5/2*a^3-a),
  x+(1/12*a^5-3/2*a^3+3*a),x+(-1/12*a^7+11/6*a^5-17/2*a^3+10*a)]
```

The factorization can also be found by evaluating the following:

b = a + 5 a^3/2 – 5 a^5/6 + a^7/24
c = 3 a – 3 a^3/2 + a^5/12
d = 10 a – 17 a^3/2 + 11 a^5/6 – a^7/12
(x–a).(x+a).(x–b).(x+b).(x–c).(x+c).(x–d).(x+d)

```
gap> a := PrimitiveElement(A);
a
gap> b := a^7/24 - 5*a^5/6 + 5*a^3/2 + a;
1/24*a^7-5/6*a^5+5/2*a^3+a
gap> c := a^5/12-3*a^3/2+3*a;
1/12*a^5-3/2*a^3+3*a
gap> d := -a^7/12 + 11*a^5/6 - 17*a^3/2 + 10*a;
-1/12*a^7+11/6*a^5-17/2*a^3+10*a
gap> (x-a)*(x-b)*(x-c)*(x-d)*(x+a)*(x+b)*(x+c)*(x+d);
x^8+(!-24)*x^6+!144*x^4+(!-288)*x^2+!144
```

This shows that the roots are $\pm a$, $\pm b$, $\pm c$, and $\pm d$, which are all expressed in terms of a. Hence, the splitting field for this polynomial is simply $\mathbb{Q}(a)$.

Since this is an eighth dimensional extension of $\mathbb{Q}$, the Galois group will have eight elements. But which group is this isomorphic to? Let us find a couple of $\mathbb{Q}$-automorphisms to find out.

By proposition 15.2, there is a $\mathbb{Q}$-automorphism f for which $f(a) = b$. Let us find this $\mathbb{Q}$-automorphism.

Homomorph[F]
Define[F[a], b]
CheckHomo[F,{a}]

```
gap> f := AlgebraHomomorphismByImagesNC(A,A,[a],[b]);
[ a ] -> [ 1/24*a^7-5/6*a^5+5/2*a^3+a ]
gap> CheckHomo(f,[a]);
true
gap> ChartHomo(f,[a,b,c,d,-a,-b,-c,-d],
  ["a","b","c","d","-a","-b","-c","-d"]);
[ a -> b, b -> -a, c -> -d, d -> c, -a -> -b, -b -> a, -c -> d,
  -d -> -c ]
```

We can have *Mathematica* draw a circle graph to find where the other seven roots are mapped to,

CircleGraph[{a, b, c, d, –a, –b, –c, –d}, F]

producing the left hand side of figure 15.3.

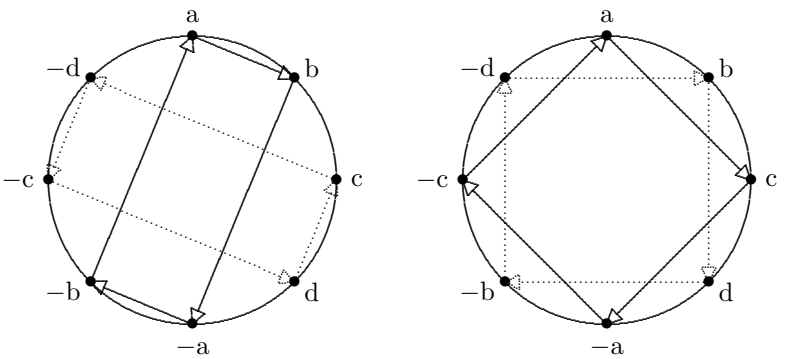


FIGURE 15.3: Two automorphisms for $x^8 - 24x^6 + 144x^4 - 288x^2 + 144$

We can express this element of the Galois group as **P[2,5,8,3,6,1,4,7]**, or (1256)(3874).

By proposition 15.2, we can also find a $\mathbb{Q}$-automorphism that sends a to c.

Homomorph[F]
Define[F[a], c]
CheckHomo[F, {a}]
CircleGraph[{a, b, c, d, –a, –b, –c, –d}, F]

This produces the circle graph on the right side of figure 15.3. In GAP, we can see where the elements b, c, and d are mapped to.

```
gap> g := AlgebraHomomorphismByImagesNC(A,A,[a],[c]);
[ a ] -> [ a^5/12-3*a^3/2+3*a ]
gap> CheckHomo(g,[a]);
true
gap> ChartHomo(g,[a,b,c,d,-a,-b,-c,-d],
  ["a","b","c","d","-a","-b","-c","-d"]);
[ a -> c, b -> d, c -> -a, d -> -b, -a -> -c, -b -> -d, -c -> a,
  -d -> b ]
```

This element of the Galois group acts like the permutation (1357)(2468) or **P[3,4,5,6,7,8,1,2]**. With these two permutations, we can see if we can generate the whole Galois group.

G = Group[{P[2,5,8,3,6,1,4,7], P[3,4,5,6,7,8,1,2]}]

```
gap> G := Group((1,2,5,6)(3,8,7,4),(1,3,5,7)(2,4,6,8));
Group([ (1,2,5,6)(3,8,7,4), (1,3,5,7)(2,4,6,8) ])
gap> Size(G);
8
gap> NumberElements := true;
true
gap> MultTable(G);

*                   |1  2  3  4  5  6  7  8
--------------------+------------------------
()                  |1  2  3  4  5  6  7  8
(1,5)(2,6)(3,7)(4,8)|2  1  4  3  6  5  8  7
(1,6,5,2)(3,4,7,8)  |3  4  2  1  7  8  6  5
(1,2,5,6)(3,8,7,4)  |4  3  1  2  8  7  5  6
(1,7,5,3)(2,8,6,4)  |5  6  8  7  2  1  3  4
(1,3,5,7)(2,4,6,8)  |6  5  7  8  1  2  4  3
(1,4,5,8)(2,7,6,3)  |7  8  5  6  4  3  2  1
(1,8,5,4)(2,3,6,7)  |8  7  6  5  3  4  1  2
```

The programs produce eight elements, so this is the entire Galois group. The multiplication table reveals that this group is isomorphic to the quaternion group Q.

Here is one more example that at first seems difficult because the splitting field is so large, but it is in fact easy to find the Galois group.

$$x^4 - x + 1.$$

In the last chapter we saw that the splitting field was 24 dimensional over $\mathbb{Q}$. We know from corollary 15.1 that the Galois group is a subgroup of S_4. But

S_4 has 24 elements, so the Galois group of $x^4 - x + 1$ must be isomorphic to S_4.

GAP has a way of determining the Galois group, up to isomorphism, for polynomials up to degree around 15 (although some polynomials of degree 14 cause a problem). Applying `GaloisType` to a polynomial produces a number, and then applying `TransitiveGroup` to this number, along with the degree of the polynomial, gives the name of the Galois group.

```
gap> x := Indeterminate(Rationals,"x");
x
gap> GaloisType(x^8-24*x^6+144*x^4-288*x^2+144);
5
gap> TransitiveGroup(8,5);
Q_8(8)
gap> GaloisType(x^5-5*x+12);
2
gap> TransitiveGroup(5,2);
D(5) = 5:2
```

In this way, we quickly redid the last two examples. However, this only gives an isomorphic group to the Galois group, instead of explicitly showing the elements of the group. Here is one last example.

```
gap> GaloisType(x^5-x+1);
5
gap> TransitiveGroup(5,5);
S5
```

Thus, GAP says that the Galois group for the polynomial $x^5 - x + 1$ is S_5.

Finally, we wish to explore a whole class of polynomials at one time. In the last chapter, we computed the splitting field of the cyclotomic polynomials, and determined that $K = \mathbb{Q}(e_n)$, where

$$e_n = e^{(2\pi i/n)} = \cos\left(\frac{2\pi}{n}\right) + i\sin\left(\frac{2\pi}{n}\right).$$

We can use proposition 14.9, along with some of the facts observed from section 13.3, to find the Galois group of the n-th cyclotomic polynomial.

PROPOSITION 15.6
Let e_n be the primitive n-th root of unity, and let $K = \mathbb{Q}(e_n)$. Then $\mathrm{Gal}_{\mathbb{Q}}(K)$ is isomorphic to a subgroup of Z_n^*.

PROOF Let $g(x) = \mathrm{Irr}_{\mathbb{Q}}(e_n, x)$. Then $g(x)$ is a factor of the n-th cyclotomic polynomial, so the roots of $g(x)$ are of the form $(e_n)^k$, where k is coprime to n. Hence, K is the splitting field of $g(x)$.

To show that $\mathrm{Gal}_{\mathbb{Q}}(K)$ is isomorphic to a subgroup of Z_n^*, note that every ϕ in $\mathrm{Gal}_{\mathbb{Q}}(K)$ is determined by where it sends e_n, and that it must send it

to one of the roots $(e_n)^k$ for some k coprime to n. Thus, there is a natural homomorphism

$$f : \mathrm{Gal}_{\mathbb{Q}}(K) \to Z_n^*$$

defined by $f(\phi) =$ (the value k for which $\phi(e_n) = (e_n)^k$). This mapping is well defined since $(e_n)^n = 1$. This mapping is a homomorphism, for if $f(\phi) = k$ and $f(\mu) = m$, then

$$(\phi \cdot \mu)(e_n) = \mu(\phi(e_n)) = \mu((e_n)^k) = (e_n)^{k \cdot m},$$

so

$$f(\phi \cdot \mu) = k \cdot m = f(\phi) \cdot f(\mu).$$

Finally, an element in the kernel of this homomorphism sends e_n to e_n, so $\mathrm{Ker}(f)$ is just the identity element of $\mathrm{Gal}_{\mathbb{Q}}(K)$. Thus, f is an isomorphism from $\mathrm{Gal}_{\mathbb{Q}}(K)$ to a subgroup of Z_n^*. □

In fact, the Galois group of the n-th cyclotomic polynomial is equal to Z_n^*, but this is harder to prove. (It requires knowing that $\Phi_n(x)$ is always irreducible.) The result given here will be sufficient for our work in the final section.

From all of these examples, we have seen a host of different groups produced as Galois groups of polynomials: S_3, Z_3, D_5, Z_5, Q, D_4, S_4, and all groups of the form Z_n^*. It is natural to ask whether *all* finite groups can be expressed as a Galois group of some polynomial in $\mathbb{Q}[x]$. This is still an open problem, known as the *inverse Galois problem.* There has been much progress made on this problem, and it is very likely to be solved soon.

While we are working with cyclotomic polynomials and n-th roots of unity, let us prove one more proposition that will be useful later on.

PROPOSITION 15.7

Let F be a finite extension of $\mathbb{Q}$ that contains the n-th roots of unity. Then if u is a root of the polynomial $f(x) = x^n - c$ for some $c \neq 0$ in F, then $K = F(u)$ is the splitting field of $f(x)$, and $\mathrm{Gal}_F(K)$ is abelian.

PROOF Since u is a root of $x^n - c$, we have that $u^n = c$. But $(e_n)^k \cdot u$ is also a root of this polynomial for all integers $k = 0, 1, 2, \cdots, n-1$, since

$$\left((e_n)^k \cdot u\right)^n = (e_n)^{k \cdot n} \cdot u^n = 1 \cdot c = c.$$

Since there are n distinct roots of the polynomial $x^n - c$ in K, the polynomial factors completely in $K[x]$. Thus, K is the splitting field of $f(x)$.

To show that $\mathrm{Gal}_F(K)$ is abelian, note that any F-automorphism is determined by where u is sent, which must be of the form $(e_n)^k \cdot u$. Thus,

if ϕ_1 and ϕ_2 are two F-automorphisms of K, where $\phi_1(u) = (e_n)^k \cdot u$ and $\phi_2(u) = (e_n)^m \cdot u$, then

$$(\phi_1 \cdot \phi_2)(u) = \phi_2(\phi_1(u)) = \phi_2((e_n)^k \cdot u) = (\phi_2(e_n))^k \phi_2(u) = (e_n)^k \cdot (e_n)^m \cdot u.$$

while

$$(\phi_2 \cdot \phi_1)(u) = \phi_1(\phi_2(u)) = \phi_1((e_n)^m \cdot u) = (\phi_1(e_n))^m \phi_1(u) = (e_n)^m \cdot (e_n)^k \cdot u.$$

Thus, $\phi_1 \cdot \phi_2 = \phi_2 \cdot \phi_1$, and so the Galois group is abelian. □

To introduce the problem of whether a fifth degree polynomial can, in general, be solved in terms of square roots, cube roots, or fifth roots, we will have *Mathematica* try to solve some polynomial equations for us. *Mathematica* can solve polynomials with the command

Solve[x^2 − x + 2 == 0]

$$\left\{\left\{x \to \frac{1}{2}(1 - i\sqrt{7})\right\}, \left\{x \to \frac{1}{2}(1 - i\sqrt{7})\right\}\right\}$$

which obviously uses the quadratic equation. Note that the "double equals" == is *Mathematica*'s way of expressing an equation. Let's try changing the x^2 to an x^3:

Solve[x^3 − x + 2 == 0]

$$\left\{\left\{x \to -\frac{\sqrt[3]{9 - \sqrt{78}}}{3^{2/3}} - \frac{1}{\sqrt[3]{3\left(9 - \sqrt{78}\right)}}\right\},\right.$$
$$\left\{x \to \frac{\left(1 + i\sqrt{3}\right)\sqrt[3]{9 - \sqrt{78}}}{23^{2/3}} + \frac{1 - i\sqrt{3}}{2\sqrt[3]{3\left(9 - \sqrt{78}\right)}}\right\},$$
$$\left.\left\{x \to \frac{\left(1 - i\sqrt{3}\right)\sqrt[3]{9 - \sqrt{78}}}{23^{2/3}} + \frac{1 + i\sqrt{3}}{2\sqrt[3]{3\left(9 - \sqrt{78}\right)}}\right\}\right\}$$

Mathematica was still able to solve this, but what a mess! The answer involves the square root of 78. Apparently *Mathematica* is using a formula that finds the roots of any cubic equation.

Let us try a forth degree equation:

Solve[x^4 − x + 2 == 0]

The answer can be expressed as

$$\left\{\left\{x \to -\frac{1}{2}\sqrt{-A - \frac{2}{\sqrt{A}}} - \frac{\sqrt{A}}{2}\right\}, \left\{x \to \frac{1}{2}\sqrt{-A - \frac{2}{\sqrt{A}}} - \frac{\sqrt{A}}{2}\right\},\right.$$
$$\left.\left\{x \to \frac{\sqrt{A}}{2} - \frac{1}{2}\sqrt{\frac{2}{\sqrt{A}} - A}\right\}, \left\{x \to \frac{1}{2}\sqrt{\frac{2}{\sqrt{A}} - A} + \frac{\sqrt{A}}{2}\right\}\right\}$$

where

$$A = \frac{\sqrt[3]{\frac{1}{2}\left(9 + i\sqrt{6063}\right)}}{3^{2/3}} + \frac{8}{\sqrt[3]{\frac{3}{2}\left(9 + i\sqrt{6063}\right)}}.$$

Once again, *Mathematica* was able to express the answer in terms of square roots and cube roots, yet this seems even more of a mess.

The equations for the cubic equation and the fourth degree equation were discovered in 1539 and 1545. [4, p. 2] The natural question is whether there is a similar formula for fifth degree polynomials. Let us try to solve a fifth degree polynomial in *Mathematica*.

Solve[x^5 − x + 2 == 0]

$$\left\{\left\{x \to \text{Root}\left[2 - \#1 + \#1^5\ \&, 1\right]\right\}, \left\{x \to \text{Root}\left[2 - \#1 + \#1^5\ \&, 2\right]\right\}, \left\{x \to \text{Root}\left[2 - \#1 + \#1^5\ \&, 3\right]\right\}, \left\{x \to \text{Root}\left[2 - \#1 + \#1^5\ \&, 4\right]\right\}, \left\{x \to \text{Root}\left[2 - \#1 + \#1^5\ \&, 5\right]\right\}\right\}$$

N[%]

$$\left\{\left\{x \to -1.26717\right\}, \left\{x \to -0.260964 - 1.17723i\right\}, \left\{x \to -0.260964 + 1.17723i\right\}, \left\{x \to 0.894548 - 0.534149i\right\}, \left\{x \to 0.894548 + 0.534149i\right\}\right\}$$

Mathematica does not know of any formula for the fifth degree polynomial, but it can find the approximate solutions. The problem is not that *Mathematica* is not smart enough to solve the equation exactly, but rather it is *impossible* to find a formula for the roots of a fifth degree polynomial in terms of square roots, cube roots, or any other roots. The reason why is based on the properties of the Galois groups. The next section will reveal how the Galois groups are related to the splitting field.

15.3 The Fundamental Theorem of Galois Theory

In this section we will clarify the relationship between subgroups of the Galois group, and the subfields of the extension field. The natural correlation is to map to each subgroup of $\text{Gal}_F(K)$ the fixed field of the subgroup. However, we ended section 15.1 with what seemed to be a bad example—$\mathbb{Q}(\sqrt[3]{2})$. The only fixed field was $\mathbb{Q}(\sqrt[3]{2})$, even though there was the obvious subfield. The way we will deal with exceptions like this one is to consider only field extensions for which the original field appears as one of the fixed fields.

DEFINITION 15.5 Let K be a finite extension of F. We say that K is a *Galois extension* if the fixed field of $\text{Gal}_F(K)$ is the field F.

Although this definition successfully rules out $\mathbb{Q}(\sqrt[3]{2})$ from being a Galois extension, we need to find a simple test for determining whether a finite extension is a Galois extension. The following proposition takes us one step in that direction.

PROPOSITION 15.8
Let F be a field, and K a Galois extension of F. If $f(x)$ is an irreducible polynomial in $F[x]$ that has at least one root in K, then $f(x)$ factors completely in K. Furthermore, $f(x)$ has no multiple roots in the field K.

PROOF Since $f(x)$ has at least one root in the field K, we will let $u_1, u_2, u_3, \ldots, u_n$ be the set of all roots of $f(x)$ in K. Consider the polynomial

$$g(x) = (x - u_1) \cdot (x - u_2) \cdot (x - u_3) \cdots (x - u_n).$$

By lemma 13.3, any automorphism in $\text{Gal}_F(K)$ extends to an automorphism on $K[x]$ with $\phi(x) = x$. Thus,

$$\phi(g(x)) = (x - \phi(u_1)) \cdot (x - \phi(u_2)) \cdot (x - \phi(u_3)) \cdots (x - \phi(u_n)).$$

By lemma 15.1, $\phi(u_1), \phi(u_2), \phi(u_3), \ldots, \phi(u_n)$ will all be roots of $f(x)$ and so this list is a permutation of the list $u_1, u_2, u_3, \ldots, u_n$. Therefore, $\phi(g(x)) = g(x)$ for all ϕ in $\text{Gal}_F(K)$.

Now, since K is a Galois extension of F, the fixed field of $\text{Gal}_F(K)$ is the field F. Thus, $g(x)$ is a polynomial in $F[x]$. Since $g(x)$ certainly divides the polynomial $f(x)$, and $f(x)$ is irreducible in $F[x]$, we have that $f(x)$ and $g(x)$ have the same degree. Thus, n is the degree of $f(x)$, and so $f(x)$ factors completely in the field K. Furthermore, $f(x)$ has no multiple roots in the field K. ∎

This proposition allows us to immediately rule out certain field extensions from being a Galois extension. Clearly $\mathbb{Q}(\sqrt[3]{2})$ is ruled out because $\mathbb{Q}(\sqrt[3]{2})$ is not a splitting field. But there are even some splitting fields that are not Galois extensions according to this proposition. Let $Z_2(t)$ be the field of rational functions in t, with coefficients in Z_2. This field can be defined in *Mathematica* by the command

InitDomain[2]

and considering rational expressions involving t. Note that there is no element whose square is equal to t.

Factor[x^2 − t]

Suppose we define a new element a that solves this equation.

Define[a^2, t]

Now $x^2 - t$ factors in $Z_2(t)(a)$ as $(x + a)(x + a)$. Note, however, that there is a *double root* in this factorization! Thus, by proposition 15.8, $Z_2(t)(a)$ is *not* a Galois extension of $Z_2(t)$.

One immediate consequence from proposition 15.8 is that a Galois extension can be written as a simple extension.

COROLLARY 15.2

Let F be a field, and let K be a Galois extension of F. Then there exists an element w of K such that $K = F(w)$.

PROOF Since K is a Galois extension of F, K is finite dimensional over F. Thus, $K = F(u_1, u_2, u_3, \ldots, u_n)$ for elements $u_1, u_2, u_3, \ldots, u_n$ in K. But the polynomials $\text{Irr}_F(u_i, x)$ all have a root in K, and so factor completely in the field K without multiple roots. Then we can use corollary 14.4 to show that there is an element w in K such that $F(w) = K$. ∎

In order to introduce the correlation between the subgroups of the Galois group and the subfields of the Galois extension, let us consider the familiar splitting field of $x^3 - 2$. Since $\sqrt[3]{2}$ and $\sqrt[3]{2}e^{2\pi i/3}$ are two roots, we can express the splitting field as $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}e^{2\pi i/3})$. The subfields of this Galois extension are $\mathbb{Q}$, $\mathbb{Q}(\sqrt[3]{2})$, $\mathbb{Q}(\sqrt[3]{2}e^{2\pi i/3})$, $\mathbb{Q}(\sqrt[3]{2}e^{4\pi i/3})$, $\mathbb{Q}(\sqrt{-3})$, and the whole field $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}e^{2\pi i/3})$. We can draw a diagram of these subfields, showing which subfields are subfields of other subfields. This is shown in figure 15.4.

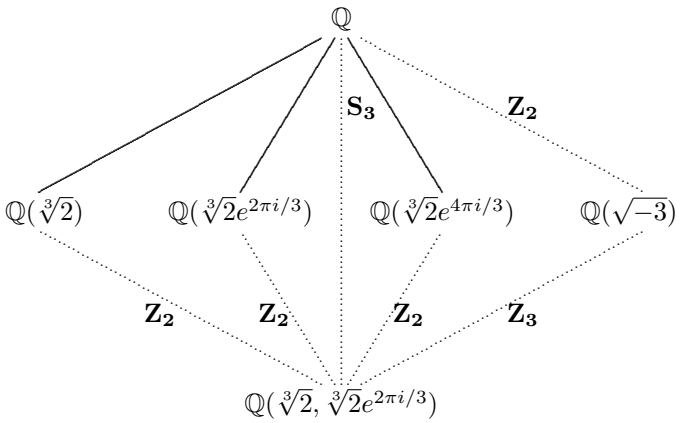


FIGURE 15.4: Subfields of $\mathbb{Q}(\sqrt[3]{2}, \sqrt[3]{2}e^{2\pi i/3})$

The dotted lines in this diagram indicate which subfields are Galois extensions of the subfield above it. Also, whenever we have a Galois extension, the corresponding Galois group is shown in boldface. For example, this diagram indicates that the splitting field of $x^3 - 2$ is a Galois extension of $\mathbb{Q}(\sqrt{-3})$. This is true by proposition 15.7, since $\mathbb{Q}(\sqrt{-3})$ contains the cube roots of unity.

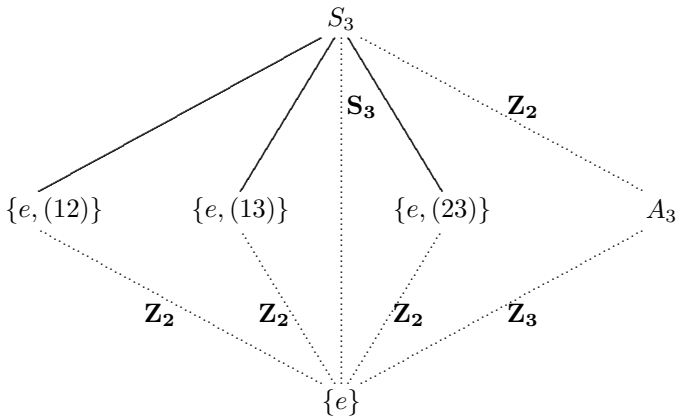


FIGURE 15.5: Subgroups of S_3

Now let us compare this figure with the subgroups of the Galois group S_3, shown in figure 15.5. Once again, we draw lines connecting two subgroups if one subgroup is contained in the other subgroup. We draw a dotted line to indicate that the smaller subgroup is a normal subgroup of the larger. Whenever the subgroup is a normal subgroup, the quotient group is indicated in boldface.

The pattern is now obvious. The two pictures are the same, except that the subfields are replaced by a subgroup of S_3. This feature of Galois extensions is the heart of Galois theory. In fact, there is a natural way that the subfields of K and the subgroups of $\mathrm{Gal}_F(K)$ are related: For each subfield E of K, we can consider $\mathrm{Gal}_E(K)$, the set of automorphisms of K that fix E. This is a subgroup of $\mathrm{Gal}_F(K)$. On the other hand, given a subgroup H of $\mathrm{Gal}_F(K)$, we can consider the fixed field $\mathrm{fix}(H)$, which is a subfield of K. To show that, indeed, the two pictures will be essentially the same, we need four steps.

1. Show that if we start with a subfield E, then form the Galois group $\mathrm{Gal}_E(K)$, and find the fixed field of this subgroup, we get back E.

2. Show that if we start with a subgroup H of $\mathrm{Gal}_F(K)$, find the fixed field, then find the Galois group of the fixed field, we get back H. These first

two steps establish a one-to-one correspondence between the subfields and the subgroups of the Galois group.

3. Show that if a subgroup N is a normal subgroup of another subgroup H, then the corresponding subfields form a Galois extension. Thus, a dotted line on the second picture corresponds to a dotted line on the first.

4. Show that if one subfield E is a Galois extension of another, L, then the corresponding Galois groups will have a normal subgroup relation. Furthermore, the quotient group of the Galois groups will be isomorphic to the Galois group of the Galois extension. Thus, a dotted line on the first picture corresponds to a dotted line on the second, and the boldface groups in the pictures will be isomorphic.

Let us begin by proving the first step.

LEMMA 15.4
Let K be a Galois extension of F, and let E be a subfield of K containing F. Then K is a Galois extension of E. That is, the fixed field of $\mathrm{Gal}_E(K)$ *is E.*

PROOF Let $H = \mathrm{Gal}_E(K)$, which is a subgroup of $\mathrm{Gal}_F(K)$. Let E_0 be the field fixed by H. Certainly E_0 contains the field E, since every automorphism in H fixes E. Suppose that u is an element of K which is not in E. Let $f(x) = \mathrm{Irr}_E(u, x)$. Since u is not in E, $f(x)$ has degree at least 2. Note that $g(x) = \mathrm{Irr}_F(u, x)$ is a polynomial in $F[x]$ for which $f(x)$ is a factor in the domain $E[x]$. Since F is a Galois field over F, $g(x)$ factors completely in K with no repeated factors. Thus, $f(x)$ also factors completely in K with no repeated factors, so there are at least two solutions to the equation $f(x) = 0$ in K. One solution is of course u, so let v be another solution. By proposition 15.2, there is an E-automorphism in H such that $\phi(u) = v$. Thus, u is not in E_0. Therefore, $E_0 = E$, and so K is a Galois extension of E. □

We are now ready to proceed to the second step.

LEMMA 15.5
Let K be a Galois extension of F. If H is a subgroup of the Galois group $\mathrm{Gal}_F(K)$*, and E is the fixed field of H, then $H = \mathrm{Gal}_E(K)$.*

PROOF Let n be the dimension of the field K over E. By lemma 15.4, K is a Galois extension of E. Thus, by corollary 15.2, there exists an element w in K such that $K = E(w)$. If $f(x) = \mathrm{Irr}_E(w, x)$, then the degree of $f(x)$ is n by corollary 14.1. Since K is a Galois extension of E, by proposition 15.8, the polynomial $f(x)$ factors completely in the field K, and there are no multiple

roots. Thus, by proposition 15.4, the number of E-automorphisms of K is the dimension of K over E, which is n.

Suppose that H contains m E-automorphisms. Let $v_1, v_2, v_3, \ldots, v_m$ be the images of w under the automorphisms in the subgroup H. That is, for each v_i there is an f in H such that $v_i = f(w)$.

Consider the polynomial

$$g(x) = (x - v_1) \cdot (x - v_2) \cdot (x - v_3) \cdots (x - v_m).$$

If ϕ is an automorphism in H, then $\phi(v_i) = \phi(f(w)) = v_j$ for some j. Also, since ϕ is one-to-one, the images of $\phi(v_1), \phi(v_2), \phi(v_3), \ldots, \phi(v_m)$ must all be distinct. Thus, each ϕ in H is a permutation on the elements $v_1, v_2, \cdots v_m$. Hence, $\phi(g(x)) = g(x)$. Since E is the fixed field $\text{fix}(H)$ of the subgroup H, we see that $g(x)$ is in $E[x]$. Thus, $f(x) = \text{Irr}_E(u, x)$ divides $g(x)$ so m is at least n. Thus,

$$|H| \leq |\text{Gal}_E(K)| = n \leq m = |H|.$$

Therefore, $H = \text{Gal}_E(K)$. □

Lemmas 15.4 and 15.5 show that there is a one-to-one correspondence between the subgroups of $\text{Gal}_F(K)$ and the subfields of K containing F. We now consider the special significance of the normal subgroups of $\text{Gal}_F(K)$.

LEMMA 15.6

Let K be a Galois extension of F, and let E be a subfield of K containing another subfield L. Suppose that $\text{Gal}_E(K)$ *is a normal subgroup of* $\text{Gal}_L(K)$. *Then every L-automorphism of K maps elements of E to elements of E. Furthermore, E is a Galois extension of L.*

PROOF First, we want to show that if u is in E, and ϕ is in $\text{Gal}_L(K)$, then $v = \phi(u)$ is in E. Since $\text{Gal}_E(K)$ is a normal subgroup of $\text{Gal}_L(K)$, for any f in $\text{Gal}_E(K)$ we have that $\psi = \phi \cdot f \cdot \phi^{-1}$ is in $\text{Gal}_E(K)$. Then $\phi \cdot f = \psi \cdot \phi$, or $f(\phi(u)) = \phi(\psi(u))$.

Since u is in E, $\psi(u) = u$, so

$$f(v) = f(\phi(u)) = \phi(\psi(u)) = \phi(u) = v.$$

Thus, v is fixed by every automorphism f in $\text{Gal}_E(K)$. By lemma 15.4, K is a Galois extension of E, so the fixed field of $\text{Gal}_E(K)$ is E. Thus, v is in E.

To show that the fixed field of $\text{Gal}_L(E)$ is L, consider an element u in E that is not in L. By lemma 15.4, K is a Galois extension of L. Since u is not in the fixed field of $\text{Gal}_L(K)$, there is an L-automorphism ϕ that moves u to another element, v. But ϕ moves all elements of E to elements of E, so we can consider the restriction of ϕ on the field E, denoted ϕ'. This is an automorphism of E, since the inverse is $(\phi^{-1})'$. Thus, there is an L-automorphism of E that

moves the element u, so the fixed field of $\mathrm{Gal}_L(E)$ is only L. Therefore, E is a Galois extension of L. □

There is only one step left to show why figures 15.4 and 15.5 are so similar.

LEMMA 15.7

Suppose that K be a Galois extension of F, and let E be a subfield of K that is also a Galois extension of a smaller subfield L. Then there exists a surjective homomorphism f from $\mathrm{Gal}_L(K)$ to $\mathrm{Gal}_L(E)$ whose kernel is $\mathrm{Gal}_E(K)$.

PROOF By lemma 15.4, K is a Galois extension of L. We begin by showing that if ϕ is an F-automorphism of K, and u is in E, then $\phi(u)$ is in E. Let $g(x) = \mathrm{Irr}_F(u, x)$. Since E is a Galois extension of L, by proposition 15.8, $g(x)$ factors completely in $E[x]$, which is of course the same factorization in $K[x]$. By lemma 15.1, $\phi(u)$ is a root of $g(x)$ in K, but all of the roots are also in E. Thus, $\phi(u)$ is in E.

Next, we define the mapping f that sends an L-automorphism of K to its restriction on the field E. We denote the restriction of ϕ on the field E by ϕ'. Since ϕ maps elements of E to elements of E, we see that ϕ' is an L-automorphism of E. However, $(\phi^{-1})'$ is also an L-automorphism of E, and $(\phi^{-1})' \cdot \phi'$ is clearly the identity mapping on E. Thus, ϕ' is an element of $\mathrm{Gal}_L(E)$.

To show that f is a homomorphism, note that

$$f(\phi_1 \cdot \phi_2) = (\phi_1 \cdot \phi_2)' = \phi_1' \cdot \phi_2' = f(\phi_1) \cdot f(\phi_2).$$

The kernel of this homomorphism is simply the L-automorphisms of K that fix the elements of E, which is of course $\mathrm{Gal}_E(K)$.

Finally, so show that this homomorphism is surjective, let ψ be an L-automorphism of E. Since K is a splitting field of E, we can use proposition 14.8 to extend ψ to an L-automorphism of K, which we will call ϕ. Then $f(\phi) = \psi$, and we have shown that f is surjective. □

Lemmas 15.4 through 15.7 explain the amazing similarity in the diagrams of the subfields, and the subgroups of the Galois group. By putting these four pieces together, we get the fundamental theorem of Galois theory.

THEOREM 15.1: The Fundamental Theorem of Galois Theory

Let K be a Galois extension of the field F. Then there is a one-to-one correspondence between the subfields of K containing F and the subgroups of $\mathrm{Gal}_F(K)$, given by mapping E to the subgroup $\mathrm{Gal}_E(K)$. The dimension of K over the subfield K is $|\mathrm{Gal}_E(K)|$. Furthermore, a subfield E is a Galois extension of L if, and only if, $\mathrm{Gal}_E(K)$ is a normal subgroup of $\mathrm{Gal}_L(K)$, in which case $\mathrm{Gal}_L(E)$ is isomorphic to $\mathrm{Gal}_L(K)/\mathrm{Gal}_E(K)$.

PROOF If $\text{Gal}_E(K) = \text{Gal}_L(K)$ for two subfields E and L of K, then by lemma 15.4, both E and L are the fixed field of the subgroup $\text{Gal}_E(K) = \text{Gal}_L(K)$, so $E = L$. Thus, the mapping $E \to \text{Gal}_E(K)$ is one-to-one. But if H is any subgroup of $\text{Gal}_F(K)$, then we can consider E to be the fixed field $\text{fix}(H)$, and by lemma 15.5 $\text{Gal}_E(K) = H$. Thus, the correspondence is also onto. Also by proposition 15.4, the dimension of K over E is $|\text{Gal}_E(K)|$, since K is a Galois extension of E.

If E is also a Galois extension of another subfield L, then by lemma 15.7 there is a surjective homomorphism from $\text{Gal}_L(K)$ to $\text{Gal}_L(E)$, whose kernel is $\text{Gal}_E(K)$. Thus, $\text{Gal}_E(K)$ is a normal subgroup of $\text{Gal}_L(K)$, and by the first isomorphism theorem (4.1), $\text{Gal}_L(E)$ is isomorphic to $\text{Gal}_L(K)/\text{Gal}_E(K)$.

Finally, suppose that $\text{Gal}_E(K)$ is a normal subgroup of $\text{Gal}_L(K)$. By lemma 15.6 E is a Galois extension of L. □

The fundamental theorem of Galois theory has many applications. With this theorem one can prove that it is impossible to trisect an angle with only a straight edge and a compass, and also that it is impossible to construct a line $\sqrt[3]{2}$ times the length of a given line. [6, p. 433] This finally puts to rest two of the three famous unsolved problems introduced by the ancient Greeks. [12, p. 109] (The last problem involves showing that π is not in an algebraic extension of $\mathbb{Q}$.) Both of these problems require a field extension of order 3, while any straight edge and compass construction involve a series of field extensions of order 2. Of course 3 does not divide any power of 2, so a field extension of dimension 3 cannot be a subfield of a field created by a sequence of extensions of order 2. The next section shows another important application of Galois theory—showing that a fifth degree equation cannot be solved in terms of radicals.

15.4 Solutions of Polynomial Equations Using Radicals

The main result of Galois theory is that one can demonstrate that it is impossible to find a formula for the solutions to a fifth degree polynomial in terms of square roots, cube roots, of fifth roots. We will spend this section exploring this problem. In fact, we will determine exactly when a polynomial can be solved in terms of radicals, and when it can't. The first step is to show that, in $\mathbb{Q}$, a Galois extension is the same thing as a splitting field.

PROPOSITION 15.9

Let E be a finite extension of $\mathbb{Q}$. If $f(x)$ is a polynomial in $E[x]$, then the splitting field of $f(x)$ is a Galois extension of E.

PROOF Let K be the splitting field of $f(x)$ in $E[x]$. If u is an element of K not in E, then $g(x) = \text{Irr}_E(u, x)$ has degree > 1. By lemma 14.5, $g(x)$ factors completely in the field K. Thus, the splitting field of $g(x)$ is contained in the field K. However, $g(x)$ is a factor of $\text{Irr}_{\mathbb{Q}}(u, x)$, which by lemma 15.3 does not have multiple roots in K. Therefore, $g(x)$ cannot have multiple roots in K, so there exists at least two roots of $g(x)$ in K. Let v be a root of $g(x)$ different from u. Then $g(x) = \text{Irr}_E(v, x)$, and so by proposition 15.2 there exists a ϕ in $\text{Gal}_E(K)$ such that $\phi(u) = v$. Thus, u is not in the fixed field of $\text{Gal}_E(K)$. Since E is obviously contained in the fixed field of $\text{Gal}_E(K)$, we find that the fixed field is E so K is a Galois extension of E. □

The next step is to give a clear definition of what it means for a polynomial to be solvable by radicals.

DEFINITION 15.6 A field K is called a *radical extension* of F if $K = F(u_1, u_2, \ldots, u_n)$, where a power of each u_i is contained in $F(u_1, u_2, \ldots, u_{i-1})$.

Here is an example of a radical extension. Suppose we considered the splitting field of the polynomial $x^4 - 8x^2 - 8x - 2$. We can have *Mathematica* solve for the roots explicitly.

```
Solve[x^4 - 8 x^2 - 8 x - 2 == 0]
```

$$\left\{ \left\{x \to -\sqrt{2} - \sqrt{2 - \sqrt{2}}\right\}, \left\{x \to -\sqrt{2} + \sqrt{2 - \sqrt{2}}\right\}, \left\{x \to \sqrt{2} - \sqrt{2 + \sqrt{2}}\right\}, \left\{x \to \sqrt{2} + \sqrt{2 + \sqrt{2}}\right\} \right\}$$

How would we express the splitting field as a radical extension? It is apparent that we first must include $\sqrt{2}$ in this field. But then it seems we need to include $\sqrt{2 + \sqrt{2}}$ and $\sqrt{2 - \sqrt{2}}$ in our field. Note, however, that the product of these two numbers is $\sqrt{2}$. Thus, all four roots are in the field $\mathbb{Q}(\sqrt{2}, \sqrt{2 + \sqrt{2}})$. This is a radical extension of $\mathbb{Q}$ of dimension 4, and the splitting field of $x^4 - 8x^2 - 8x - 2$ must be at least 4. Hence, we have found that the splitting field is a radical extension of $\mathbb{Q}$.

DEFINITION 15.7 The polynomial equation $f(x) = 0$ is said to be *solvable by radicals* if there is a radical extension of $\mathbb{Q}$ that contains the splitting field of $f(x)$.

This definition agrees with our intuitive understanding of what it means for a polynomial to be solved in terms of radicals. For example, *Mathematica*'s solution to the equation

Solve[x^3 − x + 2 == 0]

$$\left\{ \left\{ x \to -\frac{\sqrt[3]{9-\sqrt{78}}}{3^{2/3}} - \frac{1}{\sqrt[3]{3\left(9-\sqrt{78}\right)}} \right\}, \right.$$
$$\left\{ x \to \frac{\left(1+i\sqrt{3}\right)\sqrt[3]{9-\sqrt{78}}}{2\cdot 3^{2/3}} + \frac{1-i\sqrt{3}}{2\sqrt[3]{3\left(9-\sqrt{78}\right)}} \right\},$$
$$\left. \left\{ x \to \frac{\left(1-i\sqrt{3}\right)\sqrt[3]{9-\sqrt{78}}}{2\cdot 3^{2/3}} + \frac{1+i\sqrt{3}}{2\sqrt[3]{3\left(9-\sqrt{78}\right)}} \right\} \right\}$$

reveals that the splitting field is contained in radical extension

$$\mathbb{Q}\left(\sqrt{78}, \sqrt[3]{9-\sqrt{78}}, \sqrt[3]{3}, \sqrt{-3}\right).$$

This is in fact overkill, since the splitting field is at most a 6-dimensional extension of $\mathbb{Q}$, while the above radical extension may be up to a 36-dimensional extension of $\mathbb{Q}$. Yet the point is that there is *some* radical extension of $\mathbb{Q}$ that contains the roots of $x^3 - x + 2$, because the roots can be solved in terms of square roots and cube roots.

Not all radical extensions of $\mathbb{Q}$ are Galois extensions. For example, $\mathbb{Q}(\sqrt[3]{2})$ is not a Galois extension, since this extension is not the splitting field of a polynomial. In order to utilize Galois theory, we need to show that a radical extension is contained in some extension that is both a radical extension and a Galois extension.

LEMMA 15.8

Let E be a radical extension of $\mathbb{Q}$. Then E is contained in a radical extension K of $\mathbb{Q}$ such that K is a Galois extension of $\mathbb{Q}$.

PROOF Let $E = \mathbb{Q}(u_1, u_2, u_3, \ldots, u_n)$ be a radical extension of $\mathbb{Q}$. Then for every $i = 1, 2, 3, \ldots, n$, there is a k_i for which

$$(u_i)^{k_i} = v, \text{ for which } v \in \mathbb{Q}(u_1, u_2, u_3, \ldots, u_{i-1}).$$

Note that if $n = 0$, then $E = \mathbb{Q}$, and the lemma is obviously true. We will prove this by induction on n. That is, we will assume that the lemma is true for the field

$$\mathbb{Q}(u_1, u_2, u_3, \ldots, u_{n-1}).$$

That is, this field is contained in a radical extension L of $\mathbb{Q}$ that is also a Galois extension of $\mathbb{Q}$.

By corollary 15.2, there exists an element w of L such that $L = \mathbb{Q}(w)$.

Let $g(x) = \text{Irr}_{\mathbb{Q}}(w, x)$ and $p(x) = \text{Irr}_{\mathbb{Q}}(u_n, x)$. Let K be the splitting field of $g(x) \cdot p(x)$ over $\mathbb{Q}$. By proposition 15.9 K is a Galois extension of $\mathbb{Q}$. Since w is in K, L is a subfield of K. The only thing left to show is that K is a radical extension of L.

Let $v_1, v_2, v_3, \ldots, v_m$ be all of the roots of $p(x)$ in K. Since $p(x)$ is irreducible, by proposition 15.2 there is a $\mathbb{Q}$-automorphism ϕ_i that sends v_i to u_n. Since $(u_n)^k = b$ is in L, we have

$$(v_i)^k = (\phi_i(u_n))^k = \phi_i((u_n)^k) = \phi_i(b).$$

Now, L is a Galois extension of $\mathbb{Q}$, so by the fundamental theorem of Galois theory (15.1), $\text{Gal}_L(K)$ is a normal subgroup of $\text{Gal}_{\mathbb{Q}}(K)$. So by lemma 15.6 $\mathbb{Q}$-automorphisms of K map elements of L to elements of L. Thus, $\phi_i(b)$ is in L, and so $K = L(v_1, v_2, v_3, \ldots v_m)$ is a radical extension of L. □

Lemma 15.8, when combined with the definition of a polynomial solvable by radicals, tells us that if a polynomial is solvable by radicals, then the splitting field of the polynomial is contained in a field extension of $\mathbb{Q}$ that is both a radical extension and a Galois extension. What can we say about such an extension? Startlingly, the answer has a connection with the Jordan-Hölder theorem (8.2).

LEMMA 15.9

Let K be a Galois extension of $\mathbb{Q}$ which is a radical extension, and let E be a subfield of K. If E is a Galois extension of $\mathbb{Q}$, then $\text{Gal}_{\mathbb{Q}}(E)$ is a solvable group.

PROOF Since K is a radical extension of $\mathbb{Q}$, we can write

$$K = \mathbb{Q}(u_1, u_2, u_3, \ldots, u_n)$$

where some power of each u_i, $(u_i)^{k_i}$, is in $\mathbb{Q}(u_1, u_2, u_3, \ldots, u_{i-1})$.

Let m be the least common multiple of all of the k_i, and let u_0 be a primitive m-th root of unity. We would like to add u_0 in the front of the sequence of u's to get a larger field

$$M = \mathbb{Q}(u_0, u_1, u_2, u_3, \ldots, u_n).$$

Since $(u_0)^m = 1$, we see that M is still a radical extension of $\mathbb{Q}$. To show that $M = K(u_0)$ is a Galois extension of $\mathbb{Q}$, note that by corollary 15.2, $K = \mathbb{Q}(w)$ for some element w in K. If $f(x) = \text{Irr}_{\mathbb{Q}}(w, x)$, then M is the splitting field of the polynomial $f(x) \cdot (x^m - 1)$. Thus, by proposition 15.9, M is a Galois extension of $\mathbb{Q}$.

Consider the sequence of subfields

$$E_0 = \mathbb{Q}(u_0),$$

$$\begin{aligned} E_1 &= \mathbb{Q}(u_0, u_1), \\ E_2 &= \mathbb{Q}(u_0, u_1, u_2), \\ E_3 &= \mathbb{Q}(u_0, u_1, u_2, u_3), \\ &\cdots\cdots\cdots \\ E_n &= \mathbb{Q}(u_0, u_1, u_2, u_3, \ldots, u_n) = M. \end{aligned}$$

By proposition 15.7, each of these fields is a Galois extension of the previous field, since the m roots of unity were designed to be in all of these fields. Also, by proposition 15.6, E_0 is a Galois extension of $\mathbb{Q}$.

We can now apply the fundamental theorem of Galois theory (15.1). We find that $\text{Gal}_{E_i}(M)$ is a normal subgroup of $\text{Gal}_{E_{i-1}}(M)$, and the quotient group

$$\text{Gal}_{E_{i-1}}(M)/\text{Gal}_{E_i}(M)$$

is isomorphic to $\text{Gal}_{E_{i-1}}(E_i)$.

By proposition 15.7, each of these quotient groups are abelian. Also, by proposition 15.6, $\text{Gal}_{\mathbb{Q}}(E_0)$ is isomorphic to a subgroup of Z_n^*, which is abelian. Thus, the sequence of subgroups

$$\text{Gal}_{\mathbb{Q}}(M) \subseteq \text{Gal}_{E_0}(M) \subseteq \text{Gal}_{E_1}(M) \subseteq \cdots \subseteq \text{Gal}_{E_n}(M) = \{e\}$$

is a subnormal series for which all of the quotient groups are abelian. Therefore, the composition series of $\text{Gal}_{\mathbb{Q}}(M)$ will consists of only prime, cyclic factors. By the solvability theorem (8.3), $\text{Gal}_{\mathbb{Q}}(M)$ is a solvable group.

To finish the theorem, we note that E is a Galois field of $\mathbb{Q}$, so by the fundamental theorem of Galois theory (15.1), $\text{Gal}_E(M)$ is a normal subgroup of $\text{Gal}_{\mathbb{Q}}(M)$, and $\text{Gal}_{\mathbb{Q}}(E)$ is isomorphic to $\text{Gal}_{\mathbb{Q}}(M)/\text{Gal}_E(M)$. Using proposition 8.3 we see that $\text{Gal}_{\mathbb{Q}}(E)$ is solvable. □

The light is beginning to appear at the end of the tunnel. We know that any subgroup of a solvable group must be solvable. Thus, we can immediately tell whether a polynomial is solvable by radicals from its Galois group.

THEOREM 15.2: Galois' Criterion Theorem

Let $f(x)$ be a polynomial with rational coefficients. Then the equation $f(x) = 0$ is solvable by radicals only if the Galois group of $f(x)$ is a solvable group.

PROOF Suppose that $f(x)$ is a polynomial that is solvable by radicals. Let E be the splitting field of $f(x)$. By lemma 15.8, there is a field K containing E which is a Galois extension of $\mathbb{Q}$, and also is a radical extension of $\mathbb{Q}$. By proposition 15.9, E is a Galois extension of $\mathbb{Q}$. Thus, we can use lemma 15.9 to show that the Galois group of $f(x)$, $\text{Gal}_{\mathbb{Q}}(E)$ is a solvable group. □

Galois' criterion theorem is able to show us that there are some polynomials whose roots cannot be expressed in terms of square roots, cube roots, and other roots. In fact we found one of them using GAP, namely $x^5 - x + 1$.

COROLLARY 15.3

There is no formula, using only the field operations and extraction of roots, for the zeros of all fifth-degree polynomial equations.

PROOF We have already shown that the Galois group of $x^5 - x + 1$ is isomorphic to S_5. But S_5 is not solvable, since it contains the non-cyclic simple subgroup A_5. Thus, by Galois' criterion theorem (15.2) this particular equation cannot be solved with a formula involving only field operations and extraction of roots, so certainly there can be no general formula. □

Galois' criterion theorem ended the long search for a formula that finds the roots of a fifth degree polynomial. In fact, Galois' criterion theorem works the other direction as well—if the Galois group is solvable, then the polynomial *is* solvable by radicals. [2, p. 558] Since a fourth degree equation is a subgroup of S_4, which is solvable, there must be a formula for the roots of a fourth degree polynomial. The change of the structure between S_4 and S_5 is what changes the behavior of fifth degree polynomials from fourth degree polynomials.

Problems for Chapter 15

Interactive Problems

For problems **15.1** through **15.6**: Use *Mathematica* or GAP to find the Galois group of the polynomial. Determine the number of elements in the Galois group, and display a multiplication table of the subgroup of S_n isomorphic to the Galois group.

15.1 $x^4 - 2$
15.2 $x^5 - 2$
15.3 $x^5 + 15x + 12$
15.4 $x^5 + x^4 - 4x^3 - 3x^2 + 3x + 1$
15.5 $x^4 - 10x^2 + 1$
15.6 $x^8 - 108x^6 + 1548x^4 - 3888x^2 + 1296$

15.7 Use GAP or *Mathematica* to find the Galois group of $x^5 + 20x + 16$. How many elements are in the Galois group? (This may take longer than the above problems.)

Non-Interactive Problems

15.8 The Galois group $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}, \sqrt{3}))$ is given by $\{\phi_0, \phi_1, \phi_2, \phi_3\}$, where

$$\begin{array}{rcl} \phi_0(\sqrt{2}) = \sqrt{2} & \text{and} & \phi_0(\sqrt{3}) = \sqrt{3}, \\ \phi_1(\sqrt{2}) = \sqrt{2} & \text{and} & \phi_1(\sqrt{3}) = -\sqrt{3}, \\ \phi_2(\sqrt{2}) = -\sqrt{2} & \text{and} & \phi_2(\sqrt{3}) = \sqrt{3}, \\ \phi_3(\sqrt{2}) = -\sqrt{2} & \text{and} & \phi_3(\sqrt{3}) = -\sqrt{3}. \end{array}$$

Give the multiplication table for $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}, \sqrt{3}))$.

15.9 The Galois group $\mathrm{Gal}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}, \sqrt{3}))$ is given in problem 15.8. Find the five subgroups of the Galois group, and for each subgroup H find the fixed field $\mathrm{fix}(H)$ of that subgroup.

15.10 The four solutions of $x^4 - 2 = 0$ are $\sqrt[4]{2}$, $i\sqrt[4]{2}$, $-\sqrt[4]{2}$, and $-i\sqrt[4]{2}$. Thus, $K = \mathbb{Q}(\sqrt[4]{2}, i)$ is the splitting field of $x^4 - 2$. Determine the eight automorphisms of the field K, by finding where each automorphism maps the four roots.

Hint: If $\phi(r_1) = r_2$, then $\phi(-r_1) = -r_2$.

15.11 Label the three solutions of $x^3 - 3 = 0$ as $\sqrt[3]{3}$, r_2, and r_3. Determine the six automorphisms of the splitting field of $x^3 - 3$ by finding where each automorphism maps the three roots.

15.12 Find the Galois group of the field $\mathbb{Q}(\sqrt{2}, \sqrt{5})$ over $\mathbb{Q}$.

Hint: Use problem 15.8 as a model.

15.13 Find all of the subfields of the field $\mathbb{Q}(\sqrt{2}, \sqrt{5})$.

Hint: First do problem 15.12, and use the fundamental theorem of Galois theory, as was done in problem 15.9.

15.14 There are 10 subfields of the field $K = \mathbb{Q}(\sqrt[4]{2}, i)$: $\mathbb{Q}$, $\mathbb{Q}(\sqrt[4]{2}, i)$, $\mathbb{Q}(\sqrt[4]{2})$, $\mathbb{Q}(i)$, $\mathbb{Q}(i\sqrt[4]{2})$, $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(i\sqrt{2})$, $\mathbb{Q}(\sqrt{2}, i)$, $\mathbb{Q}((1+i)\sqrt[4]{2})$, and $\mathbb{Q}((1-i)\sqrt[4]{2})$. Match each of the 10 subfields with the 10 subgroups of $\mathrm{Gal}_{\mathbb{Q}}(K)$ so that each subfield is the fixed field $\mathrm{fix}(H)$ of the corresponding subgroup of $\mathrm{Gal}_{\mathbb{Q}}(K)$.

Hint: See problem 15.10 to find $\mathrm{Gal}_{\mathbb{Q}}(K)$. Next find the 10 subgroups of this group, which is isomorphic to D_4. Finding the fixed field for some of the subgroups is obvious. Can the fundamental theorem of Galois theory help with the remaining subgroups?

15.15 Find a polynomial whose Galois group is Z_6.
Hint: See proposition 15.6.

15.16 Let F be the splitting field of $\Phi_5(x) = x^4 + x^3 + x^2 + x + 1$ over $\mathbb{Q}$. Show that there is only one nontrivial subfield of F, and find this subfield.
Hint: Use proposition 15.6 to find $\mathrm{Gal}_{\mathbb{Q}}(F)$, and find that there is only one nontrivial subgroup of this group.

15.17 Prove that if a fourth degree polynomial in $\mathbb{Q}[x]$ has a Galois group isomorphic to Z_4, then the roots of the polynomial can be rearranged as r_1, r_2, r_3, and r_4 such that

$$r_1^2 r_2 + r_2^2 r_3 + r_3^2 r_4 + r_4^2 r_1$$

yields a real rational number.
Hint: There is a $\mathbb{Q}$-automorphism such that the roots map in a four-cycle: $r_1 \to r_2 \to r_3 \to r_4 \to r_1$. Note that the $\mathbb{Q}$-automorphisms fix the above expression, so the result must be in the fixed field of the Galois group.

15.18 Prove that if a fifth degree polynomial in $\mathbb{Q}[x]$ has a Galois group isomorphic to D_5, then the roots of the polynomial can be rearranged as r_1, r_2, r_3, r_4, and r_5 such that

$$r_1 r_2 + r_2 r_3 + r_3 r_4 + r_4 r_5 + r_5 r_1$$

yields a real rational number.
Hint: See the hint for problem 15.17. Note that here we must also consider a "flip" that exchanges $r_1 \leftrightarrow r_4$ and $r_2 \leftrightarrow r_3$.

15.19 Find a way similar to problem 15.17 to test whether a Galois group of a fifth degree polynomial is isomorphic to Z_5.

15.20 Find a way similar to problem 15.18 to test whether a Galois group of a fourth degree polynomial is D_4.

15.21 The roots of $x^4 - x^3 - 4x^2 + 4x + 1$ are approximately 1.827090915, 1.338261213, -1.956295201, and -0.209056927. Use trial and error to find an arrangement of these four roots such that

$$r_1^2 r_2 + r_2^2 r_3 + r_3^2 r_4 + r_4^2 r_1$$

yields an integer. (See problem 15.17.)

15.22 The roots of the equation $x^5 - 5x - 12$ are approximately 1.842085966, $0.351854083 \pm 1.709561043i$, and $-1.272897224 \pm 0.7197986815i$. Use trial and error to find an arrangement of these five roots such that

$$r_1 r_2 + r_2 r_3 + r_3 r_4 + r_4 r_5 + r_5 r_1$$

yields a real integer. (See problem 15.18.)

15.23 The roots of $x^4 - x^3 - 4x^2 + 4x + 1$ are approximately 1.827090915, 1.338261213, -1.956295201, and -0.209056927. Show that whenever a is a root, then $a^2 - 2$ is also a root. Show that, in fact, the operation $a \mapsto a^2 - 2$ permutes the four roots in a 4-cycle. Using this, prove that the Galois group must be isomorphic to Z_4.

Hint: If a is one of the roots, the splitting field is $\mathbb{Q}(a)$.

15.24 The irreducible polynomial $x^3 + x - 1$ has one real root and two complex roots. Using just this information, show that the Galois group is isomorphic to S_3.

Hint: The complex conjugate, which switches the two complex roots, is one of the $\mathbb{Q}$-automorphisms in the Galois group.

15.25 The irreducible polynomial $x^5 - 5x + 2$ has three real roots and two complex roots. Using just this information, show that the Galois group is isomorphic to S_5. (See the hint for problem 15.24.)

For problems **15.26** through **15.31**: Find a group isomorphic to the Galois group of the polynomial

15.26 $x^2 - 3$
15.27 $x^3 - 3$
15.28 $x^2 - 4$
15.29 $x^3 - 8$
15.30 $(x^2 - 2)(x^2 - 3)$
15.31 $(x - 1)^2(x - 3)^3(x^2 - 5)$

15.32 Let E be a finite extension of a field F with dimension n. Show that $|\mathrm{Gal}_F(E)| = n$ if, and only if, E is a Galois extension of F.

15.33 Let E be a finite extension of a field F, and let $\phi(x)$ be an F-automorphism in $\mathrm{Gal}_F(E)$. Suppose that $\phi(u) = u$ for some element u in E. Show that ϕ is in $\mathrm{Gal}_{F(u)}(E)$.

15.34 If E is a finite extension of $\mathbb{Q}$, and ϕ is an automorphism on E, show that ϕ is a $\mathbb{Q}$-automorphism of E.

Hint: $\phi(1) = 1$ implies that $\phi(n) = n$ for all integers n.

15.35 If E is a Galois extension of F, show that there can only be a finite number of subfields of E that contain F.

15.36 Show that if E is a Galois extension of F with dimension p, where p is a prime, prove that $\mathrm{Gal}_F(E)$ is isomorphic to Z_p.

15.37 Find, up to isomorphism, all possible Galois groups of a cubic polynomial $ax^3 + bx^2 + cx + d$.

15.38 Find, up to isomorphism, all possible Galois groups of a fourth degree polynomial $ax^4 + bx^3 + cx^2 + dx + e$.

Hint: The only subgroup of S_4 of order 8 is D_4.

15.39 Prove that if G is a group of order n that is isomorphic to a Galois group of some polynomial in $\mathbb{Q}[x]$, then G is isomorphic to a Galois group of an n-th degree polynomial in $\mathbb{Q}[x]$.

Hint: Use corollary 14.4.

This page intentionally left blank

Answers to Odd-Numbered Problems

Chapter 1

1.1) Stay = FlipRt·FlipRt, RotRt = FlipRt·FlipLft, RotLft = FlipLft·FlipRt, Spin = FlipRt·FlipLft·FlipRt.

1.3) $n = 5$, 8, or 12.

1.5) $(a.a).b \neq a.(a.b)$.

1.7) 12 steps.

1.9) $y = y \cdot e = y \cdot (x \cdot y') = (y \cdot x) \cdot y' = e \cdot y' = y'$, so $y = y'$.

1.11) 50% (18 of 36).

1.13)

	1	3	5	9	11	13
1	1	3	5	9	11	13
3	3	9	1	13	5	11
5	5	1	11	3	13	9
9	9	13	3	11	1	5
11	11	5	13	1	9	3
13	13	11	9	5	3	1

1.15)

	1	5	7	11	13	17	19	23
1	1	5	7	11	13	17	19	23
5	5	1	11	7	17	13	23	19
7	7	11	1	5	19	23	13	17
11	11	7	5	1	23	19	17	13
13	13	17	19	23	1	5	7	11
17	17	13	23	19	5	1	11	7
19	19	23	13	17	7	11	1	5
23	23	19	17	13	11	7	5	1

1.17) $100 \cdot 4 + 36 \cdot (-11) = 4$.

1.19) First find $0 \leq q \leq u \cdot v$ such that $q \equiv x (\text{Mod } u)$ and $q \equiv y (\text{Mod } v)$. Then find k so that $k \equiv q (\text{Mod } u \cdot v)$ and $k \equiv z (\text{Mod } w)$.

1.21) No, inverses would produce negatives.

1.23) If $(a \cdot b)^2 = a^2 \cdot b^2$, then $a \cdot b \cdot a \cdot b = a \cdot a \cdot b \cdot b$.

1.25) If $a^3 = e$ then $(a^{-1})^3 = e$. Furthermore, if $a \neq e$, then $a^{-1} \neq a$. So the non-identity solutions pair off, and with the identity we have an odd number of solutions.

1.27) $(n-1)((n-1)+1)/2 + n = n(n+1)/2$.

1.29) $(n-1)((n-1)+1)(2(n-1)+1)/6 + n^2 = n(n+1)(2n+1)/6$.

1.31) $(n-1)((n-1)+1)((n-1)+2)/3 + n(n+1) = n(n+1)(n+2)/3$.

1.33) If n is not prime, then $n = a \cdot b$, with $a < n$ and $b < n$.

Chapter 2

2.1) 1, 2, 4, 5, 8, 10, 11, 13, 16, 17, 19, 20.

2.3) The group has 20 elements.

2.5) $b \cdot f$ has order 15, $b \cdot f \cdot r \cdot f^2$ has order 6, $f \cdot b \cdot r$ has order 24.

2.7)

	1	5	7	11	13	17	19	23
1	1	5	7	11	13	17	19	23
5	5	1	11	7	17	13	23	19
7	7	11	1	5	19	23	13	17
11	11	7	5	1	23	19	17	13
13	13	17	19	23	1	5	7	11
17	17	13	23	19	5	1	11	7
19	19	23	13	17	7	11	1	5
23	23	19	17	13	11	7	5	1

2.9)

```
gap> f := FreeGroup("a","b");; a := f.1;; b := f.2;;
gap> g := f/[a^3, b^5, (a*b)^2];; a:= g.1;; b := g.2;;
gap> Size(g);
60
```

2.11) 3 and 5.

2.13) 40.

2.15) 288.

2.17) For $\phi(n) = 14$, either $p_i - 1$ or $p_i^{(r_i-1)}$ must be a multiple of 7 for some prime p_i. In the first case, $p_i \geq 29$, so $\phi(n) \geq 28$. In the latter case, $p_i = 7$ and $r_i \geq 2$, so $\phi(n) \geq 42$.

2.19) $b^2 \cdot a = b \cdot (a \cdot b^2) = (a \cdot b^2) \cdot b^2 = a \cdot b \cdot b^3 = a \cdot b$.

2.21) Answers will vary depending on how the elements are labeled. The group will be isomorphic to A_4.

2.23) $\{0\}$, $\{0, 2, 4, 6, 8, 10\}$, $\{0, 3, 6, 9\}$, $\{0, 4, 8\}$, $\{0, 6\}$, and the whole group.

2.25) $\{1\}$, $\{1, 2, 4, 8\}$, $\{1, 4\}$, $\{1, 4, 7, 13\}$, $\{1, 11\}$, $\{1, 14\}$, $\{1, 4, 11, 14\}$, and the whole group.

2.27) Because the corners can only rotate, every third repetition will bring the corners back to the initial state. If all 6 of the edges move, then after 6 repetitions the edges will be back in the right place, but possibly flipped. But then after 12 repetitions the edges will also be back to normal, making the order at most 12. If 5 of the edges move, then it will take 5 repetitions to get the edges into place, possibly flipped, so 10 repetitions to get the edge pieces into the right position, but then the corners may be twisted, so the order could be at most 30.

2.29) Six elements for which $x^6 = e$, three elements for which $x^3 = e$, two elements for which $x^2 = e$, so two elements of order 6. $(6 - 3 - 2$ subtracts the identity element twice.)

2.31) When $n = k$, an element is of order k if, and only if, it is a generator. If k is a divisor of n, and m is a divisor of k, then the number of solutions to $x^m = e$ will be the same in both Z_k and Z_n. Thus, computing the elements of order k in both Z_k and Z_n will give the same results.

2.33) If g is a generator, than only g and g^{-1} have finite order.

2.35) If a and b are of finite order, then $a^m = b^n = e$ for some $m > 0$ and $n > 0$. Then $(a \cdot b^{-1})^{mn} = e$, so $a \cdot b^{-1}$ is of finite order.

2.37) $(y \cdot x \cdot y^{-1})^2 = e$, but $y \cdot x \cdot y^{-1} \neq e$, so $y \cdot x \cdot y^{-1} = x$.

Chapter 3

3.1) Answers will vary.

3.3) Answers will vary.

3.5) Subgroups are $\{e\}$, with cosets $\{e\}$, $\{a\}$, $\{a^2\}$, $\{a^3\}$, $\{b\}$, $\{a \cdot b\}$, $\{a^2 \cdot b\}$, and $\{a^3 \cdot b\}$; $\{e, a^2\}$, with cosets $\{e, a^2\}$, $\{a, a^3\}$, $\{b, a^2 \cdot b\}$, and $\{a \cdot b, a^3 \cdot b\}$; $\{e, a, a^2, a^3\}$, with cosets $\{e, a, a^2, a^3\}$ and $\{b, a \cdot b, a^2 \cdot b, a^3 \cdot b\}$; $\{e, b, a^2, a^2 \cdot b\}$, with cosets $\{e, b, a^2, a^2 \cdot b\}$ and $\{a, a \cdot b, a^3, a^3 \cdot b\}$; $\{e, a \cdot b, a^2, a^3 \cdot b\}$, with cosets $\{e, a \cdot b, a^2, a^3 \cdot b\}$ and $\{a, b, a^2 \cdot b, a^3\}$; and the whole group, with one coset containing the whole group.

3.7) $\{e, b, a \cdot c, b^2, c^2, a \cdot b \cdot c, b \cdot c^2, a \cdot b^2 \cdot c, a \cdot c^3, b^2 \cdot c^2, a \cdot b \cdot c^3, a \cdot b^2 \cdot c^3\}$.

3.9) $5^{21} = 13 \pmod 7$, $7^{21} = 7 \pmod{10}$.

3.11) Since $y \in Hx$, $y = hx$ for some $h \in H$, so $Hy = H \cdot (hx) = (H \cdot h)x = Hx$.

3.13) If $n = pqr$, $\phi(n) = (p-1)(q-1)(r-1)$. If x is coprime to n, use proposition 3.1, otherwise suppose x is a multiple of p, but not a multiple of qr. Then $x^{rs} \equiv x \pmod p$, and since $rs \equiv 1 \pmod{(q-1)(r-1)}$, proposition 3.2 shows that $x^{rs} \equiv x \pmod{qr}$ as well. Finish with the Chinese remainder theorem (1.3).

3.15) Let $g_1 = x_1 \cdot y_1$ and $g_2 = x_2 \cdot y_2$ be two elements of $X \cdot Y$. Then $g_1 g_2^{-1} = (x_1 \cdot x_2^{-1}) \cdot (y_1 \cdot y_2^1) \in X \cdot Y$.

3.17) $\{0, 4, 8\}$,$\{1, 5, 9\}$,$\{2, 6, 10\}$, and $\{3, 7, 11\}$.

3.19)

	$\{0,4,8\}$	$\{1,5,9\}$	$\{2,6,10\}$	$\{3,7,11\}$
$\{0,4,8\}$	$\{0,4,8\}$	$\{1,5,9\}$	$\{2,6,10\}$	$\{3,7,11\}$
$\{1,5,9\}$	$\{1,5,9\}$	$\{2,6,10\}$	$\{3,7,11\}$	$\{0,4,8\}$
$\{2,6,10\}$	$\{2,6,10\}$	$\{3,7,11\}$	$\{0,4,8\}$	$\{1,5,9\}$
$\{3,7,11\}$	$\{3,7,11\}$	$\{0,4,8\}$	$\{1,5,9\}$	$\{2,6,10\}$

3.21)

	$\{1,4\}$	$\{2,8\}$	$\{7,13\}$	$\{11,14\}$
$\{1,4\}$	$\{1,4\}$	$\{2,8\}$	$\{7,13\}$	$\{11,14\}$
$\{2,8\}$	$\{2,8\}$	$\{1,4\}$	$\{11,14\}$	$\{7,13\}$
$\{7,13\}$	$\{7,13\}$	$\{11,14\}$	$\{1,4\}$	$\{2,8\}$
$\{11,14\}$	$\{11,14\}$	$\{7,13\}$	$\{2,8\}$	$\{1,4\}$

3.23) Since $\mathbb{Q}$ is abelian, $\mathbb{Z}$ is a normal subgroup. If $g \in \mathbb{Q}/\mathbb{Z}$, then $g = (p/q)\mathbb{Z}$ for some rational number p/q, so $g^q = p\mathbb{Z} = \mathbb{Z}$.

3.25) Let $f(x) = mx + b \in G$, and $t(x) = qx \in T$, so $f^{-1}(x) = (x-b)/m$. Then $(f \cdot t \cdot f^{-1})(x) = f^{-1}(t(f(x))) = qx + (qb - b)/m \notin T$. If $f(x) = 2x + 3$, then fT is the set of functions $k(2x+3)$, whereas Tf is the set of functions $kx + 3$.

3.27) If xN and yN are two elements in G/N, then $(xN) \cdot (yN) = x \cdot y \cdot N = y \cdot x \cdot N = (yN) \cdot (xN)$.

Chapter 4

4.1) The groups are Z_{10}:

	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9	0
2	2	3	4	5	6	7	8	9	0	1
3	3	4	5	6	7	8	9	0	1	2
4	4	5	6	7	8	9	0	1	2	3
5	5	6	7	8	9	0	1	2	3	4
6	6	7	8	9	0	1	2	3	4	5
7	7	8	9	0	1	2	3	4	5	6
8	8	9	0	1	2	3	4	5	6	7
9	9	0	1	2	3	4	5	6	7	8

and the group:

	e	a	a^2	a^3	a^4	b	$a\cdot b$	$a^2\cdot b$	$a^3\cdot b$	$a^4\cdot b$
e	e	a	a^2	a^3	a^4	b	$a\cdot b$	$a^2\cdot b$	$a^3\cdot b$	$a^4\cdot b$
a	a	a^2	a^3	a^4	e	$a\cdot b$	$a^2\cdot b$	$a^3\cdot b$	$a^4\cdot b$	b
a^2	a^2	a^3	a^4	e	a	$a^2\cdot b$	$a^3\cdot b$	$a^4\cdot b$	b	$a\cdot b$
a^3	a^3	a^4	e	a	a^2	$a^3\cdot b$	$a^4\cdot b$	b	$a\cdot b$	$a^2\cdot b$
a^4	a^4	e	a	a^2	a^3	$a^4\cdot b$	b	$a\cdot b$	$a^2\cdot b$	$a^3\cdot b$
b	b	$a^4\cdot b$	$a^3\cdot b$	$a^2\cdot b$	$a\cdot b$	e	a^4	a^3	a^2	a
$a\cdot b$	$a\cdot b$	b	$a^4\cdot b$	$a^3\cdot b$	$a^2\cdot b$	a	e	a^4	a^3	a^2
$a^2\cdot b$	$a^2\cdot b$	$a\cdot b$	b	$a^4\cdot b$	$a^3\cdot b$	a^2	a	e	a^4	a^3
$a^3\cdot b$	$a^3\cdot b$	$a^2\cdot b$	$a\cdot b$	b	$a^4\cdot b$	a^3	a^2	a	e	a^4
$a^4\cdot b$	$a^4\cdot b$	$a^3\cdot b$	$a^2\cdot b$	$a\cdot b$	b	a^4	a^3	a^2	a	e

4.3) $Z_{20}^* = \{1,3,7,9,11,13,17,19\} \approx Z_{15}^*$ with order $\{1,2,8,4,11,7,13,14\}$.
4.5) Many solutions, since b can map to either **RotLft** or **RotRt**, and a can map to **FlipLft**, **FlipRt**, or **Spin**. Any of these combinations will work.
4.7) If $f(x) = a$ and $f(y) = b$, then $f^{-1}(a \cdot b) = x \cdot y = f^{-1}(a) \cdot f^{-1}(b)$.
4.9) $1 \mapsto 0$, $-1 \mapsto 2$, $\pm i$ can go to either 1 or 3.
4.11) $Z_6 = \{0,1,2,3,4,5\} \approx Z_9^*$ with order $\{1,2,4,8,7,5\}$.
4.13) $Z_6 = \{0,1,2,3,4,5\} \approx Z_{18}^*$ with order $\{1,5,7,17,13,11\}$.
4.15) $Z_{10} = \{0,1,2,3,\ldots,9\} \approx Z_{22}^*$ with order $\{1,7,5,13,3,21,15,17,9,19\}$.
4.17) $Z_{12} = \{0,1,2,3,4,5,6,7,8,9,10,11\} \approx Z_{26}^*$, using the arrangement $\{1,7,23,31,9,11,25,19,3,21,17,15\}$.
4.19) Not true if G is not abelian.
4.21) $x \cdot (H \cdot N) = (x \cdot H) \cdot N = (H \cdot x) \cdot N = H \cdot (x \cdot N) = (H \cdot N) \cdot x$.
4.23) If g is a generator of G, and $x \in \mathrm{Im}(\phi)$, then $x = \phi(g^n) = (\phi(g))^n$ for some n, and hence $\phi(g)$ generates $\mathrm{Im}(\phi)$.
4.25) $\phi(x \cdot y) = \phi(x + y) = 2(x + y) = 2x + 2y = \phi(x) + \phi(y) = \phi(x) \cdot \phi(y)$, since $\cdot$ is addition in this group.
4.27) $\phi(x \cdot y) = \phi(x + y) = x + y + 3$, but $\phi(x) \cdot \phi(y) = \phi(x) + \phi(y) = (x + 3) + (y + 3) = x + y + 6$.
4.29) $\phi(x \cdot y) = 2(x \cdot y) = 2xy$, but $\phi(x) \cdot \phi(y) = (2x) \cdot (2y) = 4xy$.

4.31) $\phi(x \cdot y) = \phi(x + y) = e^{x+y} = e^x \times e^x = \phi(x) \cdot \phi(y)$. Image is the positive real numbers.

4.33) $\phi(f \cdot g) = \phi(f(t) + g(t)) = f(3) + g(3) = \phi(f) + \phi(g) = \phi(f) \cdot \phi(g)$. The kernel is the set of polynomials with 3 as a root, hence $t - 3$ is a factor.

4.35) $\phi(1) = 1$, $\phi(7) = 13$, $\phi(11) = 1$, $\phi(13) = 7$, $\phi(17) = 13$, $\phi(19) = 19$, $\phi(23) = 7$, $\phi(29) = 19$.

4.37) $\phi(x \cdot y) = [x \cdot y \ (\text{Mod } n)] \text{ Mod } k = x \cdot y \text{ Mod } k = \phi(x) \cdot \phi(y)$. The kernel is the multiples of k, so there are n/k elements in the kernel.

4.39) Ten homomorphisms, one sending all elements to e, three sending $\{1, 3\}$ to e, $\{5, 7\}$ to a, $a \cdot b$, or $a \cdot b^2$ respectively, three sending $\{1, 5\}$ to e, $\{3, 7\}$ to a, $a \cdot b$, or $a \cdot b^2$ respectively, and three sending $\{1, 7\}$ to e, $\{3, 5\}$ to a, $a \cdot b$, or $a \cdot b^2$ respectively.

4.41) Since $\{0, 2, 4\}$ and $\{0, 3\}$ are normal subgroups of Z_6, $\phi^{-1}(\{0, 2, 4\})$ and $\phi^{-1}(\{0, 3\})$ are normal subgroups of G.

Chapter 5

5.1) 3-cycle example: $(123)(324) = (143)$; but 4-cycles are odd.

5.3) By using the ordering $\{1, 5, 7, 11, 13, 17, 19, 23\}$, we get the permutations $()$, $(12)(34)(56)(78)$, $(13)(24)(57)(68)$, $(14)(23)(58)(67)$, $(15)(26)(37)(48)$, $(16)(25)(38)(47)$, $(17)(28)(35)(46)$, $(18)(27)(36)(45)$.

5.5) $P[7, 6, 4, 1, 2, 5, 3] = (1734)(265)$ and $P[4, 6, 7, 3, 2, 5, 1] = (1437)(265)$.

5.7) $\left(\begin{smallmatrix} 1&2&3&4&5&6 \\ 3&4&5&1&2&6 \end{smallmatrix}\right)$.

5.9) $\left(\begin{smallmatrix} 1&2&3&4 \\ 2&3&4&1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1&2&3&4 \\ 2&4&1&3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1&2&3&4 \\ 3&4&2&1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1&2&3&4 \\ 3&1&4&2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1&2&3&4 \\ 4&3&1&2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1&2&3&4 \\ 4&1&2&3 \end{smallmatrix}\right)$.

5.11) $x = \left(\begin{smallmatrix} 1&2&3&4 \\ 2&4&1&3 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} 1&2&3&4 \\ 3&2&1&4 \end{smallmatrix}\right)$, or $\left(\begin{smallmatrix} 1&2&3&4 \\ 4&3&1&2 \end{smallmatrix}\right)$.

5.13) $(16453)(27)$.

5.15) $(1568)(37)$.

5.17) 6 and 12.

5.19) $(12345)(678) \in A_8$, since this is an even permutation.

5.21) $\left\{\left(\begin{smallmatrix} 1&2&3&4 \\ 1&2&3&4 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1&2&3&4 \\ 2&1&4&3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1&2&3&4 \\ 3&4&1&2 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1&2&3&4 \\ 4&3&2&1 \end{smallmatrix}\right)\right\}$.

5.23) -1 must map to a product of two transpositions, like $(12)(34)$. Then $\pm i$, $\pm j$, and $\pm k$ map to one of (1324), (1423), $(1324)(56)$, $(1324)(57)$, $(1324)(67)$, $(1423)(56)$, $(1423)(57)$, or $(1423)(67)$. But no combination of these allows $i \cdot j = k$.

5.25) Technically, $\left(\begin{smallmatrix} 1&2&3&4&5 \\ 2&1&4&3&5 \end{smallmatrix}\right) \in S_5$, and $\left(\begin{smallmatrix} 1&2&3&4 \\ 2&1&4&3 \end{smallmatrix}\right) \in S_4$, which are totally different groups. However, there is a natural mapping from S_4 to S_5 that allows us to consider elements of S_4 to also be in S_5.

5.27) If ϕ_1 and ϕ_2 only move a finite number of integers, then $\phi_1 \cdot \phi_2^{-1}$ will move a finite number of integers. Also, if n is the largest integer that ϕ_1 moves, then $\phi_1 \in S_n$ in the sense of problem 5.25, so $S_\Omega \subseteq \bigcup_{n=1}^{\infty} S_n \subseteq S_\Omega$.

5.29) If ϕ_1 has $|\phi_1(x) - x| < M$ for all x, and ϕ_2 has $|\phi_2(x) - x| < N$ for all x, then $|\phi_2^{-1}(y) - y| < N$ for all $y = \phi^{-1}(x)$, and $|\phi_1(\phi_2^{-1}(x)) - x| <$

$M + N$ for all x. Examples: $(12)(34)(56)\ldots(2n-1\ \ 2n)\ldots \in G$, but $\notin S_\Omega$; $(12)(46)(9\ 12)(16\ 20)\ldots(n^2\ \ n^2+n)\ldots \in S_\infty$, but $\notin G$.

5.31) Applying corollary 5.2: $p \cdot m$ divides $m! \cdot |N|$, so p divides $|N|$, hence $H = N$, and H is normal.

5.33) 144.

5.35) Let H be the subgroup generated by the n-cycle $\phi = (123\ldots n)$. Then ϕ^{j-i} will map i to j.

5.37) If $\phi = (i_1\ i_2\ i_3\ \ldots i_r)$ and $f = (j_1\ j_2\ j_3\ \ldots j_s)$, then $x^{-1} \cdot \phi \cdot x = (x(i_1)\ x(i_2)\ x(i_3)\ \ldots x(i_r))$, and $x^{-1} \cdot f \cdot x = (x(j_1)\ x(j_2)\ x(j_3)\ \ldots x(j_s))$.

Chapter 6

6.1) $Z_2 \times Z_6$ has three elements of order 2, whereas Z_{12} has only one element of order 2.

6.3) 55.

6.5) Eight automorphisms: $\{1,2,4,7,8,11,13,14\} \mapsto \{1,2,4,7,8,11,13,14\}$, {1, 2, 4, 13, 8, 14, 7, 11}, {1, 7, 4, 2, 13, 11, 8, 14}, {1, 7, 4, 8, 13, 14, 2, 11}, {1, 8, 4, 7, 2, 14, 13, 11}, {1, 8, 4, 13, 2, 11, 7, 14}, {1, 13, 4, 2, 7, 14, 8, 11}, or {1, 13, 4, 8, 7, 11, 2, 14}.

6.7) There are 20 automorphisms, generated by $f(a) = a$, $f(b) = b^2$, and $g(a) = a \cdot b$, $g(b) = b$.

6.9) A nontrivial homomorphism from Z_8^* to $\mathrm{Aut}(Z_8^*) \approx S_3$ must be two-to-one, and send two of the elements to a 2-cycle. Proposition 6.7 shows that it does not matter which 2-cycle, and since the non-identity elements of Z_8^* are essentially equivalent, there is isomorphically only one $Z_8^* \ltimes Z_8^* \approx Z_2 \times D_4$.

6.11) $D_6 \approx S_3 \times Z_2$.

6.13) $\{(0,0),(0,1),(1,0),(1,1),(2,0),(2,1)\} \mapsto \{0,3,4,1,2,5\}$.

6.15)

	(0,1)	(0,3)	(0,5)	(0,7)	(1,1)	(1,3)	(1,5)	(1,7)	(2,1)	(2,3)	(2,5)	(2,7)
(0,1)	(0,1)	(0,3)	(0,5)	(0,7)	(1,1)	(1,3)	(1,5)	(1,7)	(2,1)	(2,3)	(2,5)	(2,7)
(0,3)	(0,3)	(0,1)	(0,7)	(0,5)	(1,3)	(1,1)	(1,7)	(1,5)	(2,3)	(2,1)	(2,7)	(2,5)
(0,5)	(0,5)	(0,7)	(0,1)	(0,3)	(1,5)	(1,7)	(1,1)	(1,3)	(2,5)	(2,7)	(2,1)	(2,3)
(0,7)	(0,7)	(0,5)	(0,3)	(0,1)	(1,7)	(1,5)	(1,3)	(1,1)	(1,7)	(1,5)	(1,3)	(1,1)
(1,1)	(1,1)	(1,3)	(1,5)	(1,7)	(2,1)	(2,3)	(2,5)	(2,7)	(0,1)	(0,3)	(0,5)	(0,7)
(1,3)	(1,3)	(1,1)	(1,7)	(1,5)	(2,3)	(2,1)	(2,7)	(2,5)	(0,3)	(0,1)	(0,7)	(0,5)
(1,5)	(1,5)	(1,7)	(1,1)	(1,3)	(2,5)	(2,7)	(2,1)	(2,3)	(0,5)	(0,7)	(0,1)	(0,3)
(1,7)	(1,7)	(1,5)	(1,3)	(1,1)	(2,7)	(2,5)	(2,3)	(2,1)	(0,7)	(0,5)	(0,3)	(0,1)
(2,1)	(2,1)	(2,3)	(2,5)	(2,7)	(0,1)	(0,3)	(0,5)	(0,7)	(1,1)	(1,3)	(1,5)	(1,7)
(2,3)	(2,3)	(2,1)	(2,7)	(2,5)	(0,3)	(0,1)	(0,7)	(0,5)	(1,3)	(1,1)	(1,7)	(1,5)
(2,5)	(2,5)	(2,7)	(2,1)	(2,3)	(0,5)	(0,7)	(0,1)	(0,3)	(1,5)	(1,7)	(1,1)	(1,3)
(2,7)	(2,7)	(2,5)	(2,3)	(2,1)	(0,7)	(0,5)	(0,3)	(0,1)	(1,7)	(1,5)	(1,3)	(1,1)

6.17) Since $x^n = e$ for all $x \in Z_n \times Z_n$, we see that $Z_n \times Z_n$ is not cyclic.

6.19) Only Z_{210}.

6.21) Four abelian groups of order 36.

6.23) 100.

6.25) $Z_4 \times Z_2 \times Z_5$.

6.27) Note that any automorphism must fix the identity element, leaving $n-1$ elements.
6.29) If $\mathrm{Aut}(G)$ is cyclic, then so is $\mathrm{Inn}(G)$ with a generator $x \mapsto g^{-1}xg$. For each $y \in G$, $y^{-1}xy = g^{-n}xg^n$ for some n, plugging in $x = g$ yields $y^{-1}gy = g$, or $gy = yg$. Since $gy = gy$ for all y, $\mathrm{Inn}(G) \approx \{e\}$, and G is abelian.
6.31) Since Z_3^* and Z_4^* both have two elements, we can pick $G = Z_3$ and $M = Z_4$.
6.33) A nontrivial mapping from Z_3 to $\mathrm{Aut}(Z_8^*)$ maps 1 to a 3-cycle, which by proposition 6.7 doesn't matter which. $Z_3 \ltimes Z_8^* \approx A_4$.
6.35) Since $\mathrm{Aut}(\mathbb{Z}) \approx Z_2$, we see that $\phi_1(x) = -x$. So $(a, x) \cdot (b, y) = (a + b, x + y)$ when b is even, but $(a, x) \cdot (b, y) = (a + b, y - x)$ when a is odd.

Chapter 7

7.1) Center $= \{e, b^3\}$, Quotient group $D_6/Z(D_6) \approx S_3$.
7.3) $\{e\}$, $\{e, b^3\}$, $\{e, b^2, b^4\}$, $\{e, b, b^2, b^3, b^4, b^5\}$, $\{e, b^2, b^4, a, a{\cdot}b^2, a{\cdot}b^4\}$, $\{e, b^2, b^4, a \cdot b, a \cdot b^3, a \cdot b^5\}$, and D_6.
7.5) Five subgroups of order 4, one subgroup of order 5.
7.7) $\{e, b^2\}$.
7.9) $N_{D_4}(\{e\}) = N_{D_4}(\{b^2\}) = D_4$, $N_{D_4}(\{b\}) = N_{D_4}(\{b^3\}) = \{e, b, b^2, b^3\}$, $N_{D_4}(\{a\}) = N_{D_4}(\{a \cdot b^2\}) = \{e, a, b^2, a \cdot b^2\}$, $N_{D_4}(\{a \cdot b\}) = N_{D_4}(\{a \cdot b^3\}) = \{e, a \cdot b, b^2, a \cdot b^3\}$.
7.11) Yes, if x and y are in the center, then $x \cdot y = y \cdot x$.
7.13) $\{e\}$, $\{b^2\}$, $\{b, b^3\}$, $\{a, a \cdot b^2\}$, and $\{a \cdot b, a \cdot b^3\}$.
7.15) $\{e\}$, $\{a, a \cdot b, a \cdot b^2, a \cdot b^3, a \cdot b^4\}$, $\{b, b^4\}$, and $\{b^2, b^3\}$.
7.17) $x \in Z(G) \Leftrightarrow x \cdot y = y \cdot x$ for all $y \in G \Leftrightarrow x \in N_G(\{y\})$ for all $y \in G$.
7.19) If N is a nontrivial normal subgroup, $|N| \geq 13$, so $|N| = 30$, 20, or 15 (divisors of 60). $|N| \neq 15$, so $|N|$ is even, hence classes of size 1 and 15 are in N. Since $|N| \geq 28$, $|N| = 30$, but there is no class of size 14.
7.21) $|N| \geq 41$, so $|N| = 180$, 120, 90, 72, 60, or 45 (divisors of 360). $|N| \neq 45$, so $|N|$ is even, hence classes of size 1 and 45 are in N, making $|N| \geq 86$. 10 divides $|N|$, so both classes of order 72 are in N, making $|N| \geq 230$.
7.23) $|N| \geq 56$, so $|N| = 330$, 220, 165, 132, 110, 66, or 60 (divisors of 660). $|N| \neq 60$, so 11 divides $|N|$, hence both classes of size 60 are in N, making $|N| \geq 176$. Five divides $|N|$, so both classes of order 132 are in N, making $|N| \geq 385$.
7.25) $|N| \geq 316$, so $|N| = 10080$, 6720, 5040, 4032, 3360, 2880, 2520, 2240, 2016, 1680, 1440, 1344, 1260, 1120, 1008, 960, 840, 720, 672, 630, 576, 560, 504, 480, 448, 420, 360, 336, or 320 (divisors of 20160). $|N|$ is even, so classes of size 1 and 315 are in N, making $|N| \geq 1576$. $|N| \neq 2240$, so $|N|$ is a multiple of 3, so the class of size 2240 is in N, making $|N| \geq 3816$. Seven divides $|N|$, so both classes of size 2880 are in N, making $|N| \geq 9576$. Five divides $|N|$, so both classes of size 4032 are in N, making $|N| \geq 16380$. A_8 has a conjugacy class of size 112 (all 3-cycles).
7.27) 20160 elements, same as A_8 and $L_3(4)$ from problem 7.25. This group is in fact isomorphic to A_8.

7.29) Let K be any p-Sylow subgroup of size p^n, and divide G into families, where u and v are related if $u = h \cdot v \cdot k$ for $h \in H$ and $k \in K$. Then $|G| = p^n \cdot m = \sum p^i \cdot p^n / |H \cap (u_j \cdot K \cdot u_j^{-1})|$, so $|H \cap (u_j \cdot K \cdot u_j^{-1})| = p^i$ for some j, meaning that H is completely contained in a p-Sylow subgroup.
7.31) There are either one or eight 7-Sylow subgroups. If not unique, there are 48 elements of order 7, leaving 8 elements for a unique 2-Sylow subgroup.
7.33) There is only one 3-Sylow subgroup H, and only one 11-Sylow subgroup N, so both are normal, and $G \approx H \times N$. Thus, $G \approx Z_{99}$ or $Z_3 \times Z_3 \times Z_{11}$.
7.35) There is only one 17-Sylow subgroup N, 1 or 51 5-Sylow subgroups, and 1 or 85 3-Sylow subgroups. Either a 3-Sylow subgroup H or 5-Sylow subgroup K is normal, so $H \cdot K$ is a subgroup of order 15 $\approx Z_{15}$. Then $G \approx Z_{15} \times Z_{17} \approx Z_{255}$, or $G \approx Z_{17} \ltimes_\phi Z_{15}$. But there is no nontrivial homomorphism between Z_{15} and Z_{17}^*.
7.37) Factors of $|G|$ are 1, p, p^2, q, pq, p^2q. There are either 1 or q p-Sylow subgroups, and either 1, p, or p^2 q-Sylow subgroups. If neither are unique, $q \equiv 1$ (Mod p), implying $p < q$, so $p^2 \equiv 1$ (Mod q). Then we have $p^2(q-1)$ elements of order q, leaving only p^2 elements for a normal p-Sylow subgroup.
7.39) Only cases not covered by problems 7.36 through 7.38 or proposition 7.8 are 30, 36, 42, and 48. If $G = 30$, there aren't enough elements for both 10 3-Sylow subgroups and 6 5-Sylow subgroups. If $G = 36$, there is a 3-Sylow subgroup of order 9, and applying corollary 5.2 gives a normal subgroup of size 3 or 9. If $G = 42$, there is only one 7-Sylow subgroup. If $G = 48$, there is a 2-Sylow subgroup of order 16, and applying corollary 5.2 gives a normal subgroup of size 8 or 16.

Chapter 8

8.1) $Q \supseteq \{1, -1\} \supseteq \{1\}$. For compositions series, add $\{1, -1, i, -i\}$.
8.3) $G' \approx Q$, which is a normal subgroup of G, and there is a 3-Sylow subgroup H for which $H \cdot G' = G$. Hence, G is isomorphic to a semi-direct product $Q \ltimes_\phi Z_3$, and since $\mathrm{Aut}(Q) \approx S_4$, Z_3 must map to a 3-cycle in S_4, but all 3-cycles are conjugate, so there is only one possible semi-direct product $G \approx Q \ltimes Z_3$.
8.5) $B \supseteq \{1, L, P, T, I, M, Q, U\} \supseteq \{1, L, P, T\} \supseteq \{1\}$; if $a = P$, $b = I$, and $c = J$, then $a^4 = 1$, $b^2 = a^2$, $c^2 = a^2$, $b \cdot a = a \cdot b$, $c \cdot a = a \cdot c$, $c \cdot b = a \cdot a \cdot b \cdot c$.
8.7) $D \supseteq \{1, L, M, N, O, P, Q, R\} \supseteq \{1\}$; if $a = L$ and $b = S$, then $a^8 = 1$, $b^2 = a^4$, $b \cdot a = a^7 \cdot b$.
8.9) $A_{1,1} = A_{1,2} = B_{1,1} = Z_{12}$, $A_{2,1} = \{0, 6\}$, $A_{2,2} = B_{1,3} = \{0\}$, $B_{1,2} = \{0, 2, 4, 6, 8, 10\}$. The arrows show the isomorphisms $Z_{12}/Z_{12} \approx Z_{12}/Z_{12}$, $Z_{12}/Z_{12} \approx \{0, 2, 4, 6, 8, 10\}/\{0, 2, 4, 6, 8, 10\}$, $Z_{12}/\{0, 3, 6, 9\} \approx \{0, 4, 8\}/\{0\}$, $\{0, 3, 6, 9\}/\{0, 6\} \approx Z_{12}/\{0, 2, 4, 6, 8, 10\}$, $\{0, 6\}/\{0\} \approx \{0, 2, 4, 6, 8\}/\{0, 4, 8\}$, $\{0\}/\{0\} \approx \{0\}/\{0\}$.
8.11) $Z_{24}^* \supseteq \{1, 5, 7, 11\} \supseteq \{1, 5\} \supseteq \{1\}$.
8.13) $Z_{12} \times Z_{18} \supseteq \{0, 3, 6, 9\} \times Z_{18} \supseteq \{0, 6\} \times Z_{18} \subseteq \{0\} \times Z_{18} \supseteq \{0\} \times \{0, 3, 6, 9, 12, 15\} \supseteq \{0\} \times \{0, 9\} \supseteq \{0\} \times \{0\}$.
8.15) $D_4 \subseteq \{e, b, b^2, b^3\} \subseteq \{e, b^2\} \subseteq \{e\}$.

8.17) $D_6 \subseteq \{e, b, b^2, b^3, b^4, b^5\} \subseteq \{e, b^3\} \subseteq \{e\}$.
8.19) A_4 and $\{(), (12)(34), (13)(24), (14)(23)\}$ must be in the series, and then we have three choices, $\{(), (12)(34)\}$, $\{(), (13)(24)\}$, or $\{(), (14)(23)\}$ for the next term in the series.
8.21) $Z_4 \supseteq \{0, 2\} \supseteq \{0\}$, and $Z_8^* \supseteq \{1, 3\} \supseteq \{1\}$.
8.23) $A_5 \times A_5 \supseteq \{e\} \times A_5 \supseteq \{e\} \times \{e\}$.
8.25) $G' = \{\phi(x) = x + c \mid c \in \mathbb{R}\}$, both $G' \approx \mathbb{R}$ and $G/G' \approx \mathbb{R}$.
8.27) Z_{16}, $Z_8 \times Z_2$, $Z_4 \times Z_4$, $Z_4 \times Z_2 \times Z_2$, and $Z_2 \times Z_2 \times Z_2 \times Z_2$ are the only groups that are abelian, and by the fundamental theorem of finite abelian groups (6.2) these are all non-isomorphic. $Z_2 \times D_8$ has 11 elements of order 2, D_{16} has 9, G from section 6.4 has 5, and D from problem 8.7 has only 1 element of order 2. B from problem 8.5 and C from problem 8.6 both have 7 elements of order 2, but B has only 2 elements along the diagonal, whereas C has 4. Finally, M from section 6.4, $Z_2 \times Q$, and $Z_4 \ltimes Z_4$ have 3 elements of order 2, but $Z_2 \times Q$ has only 2 elements along the diagonal, M has 4 elements along the diagonal, and $Z_4 \ltimes Z_4$ has 3 elements along the diagonal.
8.29) By problem 1.22, G is abelian, hence solvable. But for G/N to be cyclic, then G/N would be of order 2, and N would have the same properties. Thus, a polycyclic series would not reach $\{e\}$ in a finite number of steps.
8.31) $(D_4)' = \{e, b^2\}$, $(D_4)'' = \{e\}$.
8.33) $Q' = \{1, -1\}$, $Q'' = \{1\}$.
8.35) If $G = S_4$, then $G_1 = [S_4, S_4] = A_4$, but $G_2 = [S_4, A_4] = A_4$, so G_n will never go to $\{e\}$.
8.37) Since all of the A_i and B_j are normal subgroups of G, then $A_{i,j} = (A_{i-1} \cap B_j) \cdot A_i$ and $B_{j,i} = (B_{j-1} \cap A_i) \cdot B_j$ are normal subgroups of G using problem 4.21.
8.39) If $G = N_0 \supseteq N_1 \supseteq \cdots \supseteq N_k = \{e\}$ is a chief series, then $G' \subseteq N_1$ by lemma 8.3. Define $G_1 = G'$, $G_2 = [G', G_1]$, $G_3 = [G', G_2]$, ..., and suppose by induction that $G_i \subseteq N_i$. We must show that $[G', N_i] \subseteq N_{i+1}$, since this would indicate that $G_k = \{e\}$. Since N_i/N_{i+1} is cyclic, there is a generator nN_{i+1}. For $x, y \in G$, we have $x \cdot n \cdot x^{-1} N_{i+1} = n^q N_{i+1}$ for some q, and $y \cdot n \cdot y^{-1} N_{i+1} = n^r N_{i+1}$ for some r. Then $y^{-1} \cdot x^{-1} \cdot y \cdot x \cdot n^{-1} \cdot x^{-1} \cdot y^{-1} \cdot x \cdot y \cdot nN_{i+1} = N_{i+1}$, so $[x^{-1} \cdot y^{-1} \cdot x \cdot y, n] \in N_{i+1}$. Thus, $[G', N_i] \subseteq N_{i+1}$.
8.41) $|S_8 \ltimes_\phi (Z_3 \times Z_3 \times Z_3 \times Z_3 \times Z_3 \times Z_3 \times Z_3)| = 88179840$.

Chapter 9

9.1) $y = (4/3)^{10}(3/4)^{(1/x)}$.
9.3)

```
gap> InitRing("a","b","c");
gap> DefineRing("R",[2,2,2],[[a,a,c],[b,b,a+b+c],[a,a,c]]);
```

9.5)

```
gap> InitRing("e","a","b");
gap> DefineRing("T8",[2,2,2],[[e,a,b],[a,a,a],[b,b,b]]);
```

9.7) If $p^3/q^3 = 2$ with p and q coprime, then $2|p$, but replacing $p = 2r$ shows $2|q$ too.

9.9) Given x and y, choose any irrational z, and find a rational q between $x-z$ and $y-z$. Then $q+z$ is irrational by problem 9.8.
9.11) $x^2 = 13+2\sqrt{6}$, and $\sqrt{6}$ is irrational, so x^2 is too. If x were rational, then x^2 would be rational.
9.13) $(a_1-b_1i-c_1j-d_1k)+(a_2-b_2i-c_2j-d_2k) = (a_1+a_2)-(b_1+b_2)i-(c_1+c_2)j(d_1+d_2)k$.
9.15) $(a+bi+cj+dk)\cdot(a-bi-cj-dk) = a^2+b^2+c^2+d^2$. Replace x with $\overline{x}$ to get the other half.
9.17) $(x+1)\cdot(x-1) = x^2+x-x-1 = x^2-1$.
9.19) If $a = x_1+y_1\sqrt{2}$ and $b = x_2+y_2\sqrt{2}$, then $a-b = (x_1-x_2)+(y_1-y_2)\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$, $a\cdot b = (x_1x_2+2y_1y_2)+(x_1y_2+x_2y_1)\sqrt{2} \in \mathbb{Z}[\sqrt{2}]$.
9.21) Both $x\cdot a = x$ and $x\cdot b = x$ for all x in the ring, but there is no r for which $r\cdot c = c$, since $r\cdot c = 0$.
9.23) Since $n(-x)+nx = n(-x+x) = 0$, we have $n(-x) = -nx$.
9.25) Since G is an abelian group, we only need to check the associate law and the two distributive laws. But these are both trivial, since both sides would evaluate to 0.
9.27) $x = x^2 = (-x)^2 = -x$.
9.29) $\oplus$ and $\otimes$ are both closed, and both are clearly commutative. $(x\oplus y)\oplus z = x+y+z-2 = x\oplus(y\oplus z)$, $x\oplus 1 = 1\oplus x = x$ so 1 is the additive identity. $x\oplus(2-x) = 1$, so $2-x$ is the additive inverse. $(x\otimes y)\otimes z = x+y+z-xy-xz-yz+xyz = x\otimes(y\otimes z)$, $x\otimes(y\oplus z) = 2x+y+z-xy-xz-1 = (x\otimes y)\oplus(x\otimes z)$.
9.31) Obviously 0 and 1 satisfy $a^2 = a$. If $a \neq 0$, then a^{-1} exists, and $a = a^2\cdot a^{-1} = a\cdot a^{-1} = 1$.
9.33) First show $\binom{n-1}{i-1}+\binom{n-1}{i} = \binom{n}{i}$. Then

$$\begin{aligned}
&(x+y)\cdot\left(x^{n-1}+\binom{n-1}{1}x^{n-2}y+\binom{n-1}{2}x^{n-3}y^2+\cdots+\binom{n-1}{n-1}y^{n-1}\right)\\
&= x^n+\left[1+\binom{n-1}{1}\right]x^{n-1}y+\left[\binom{n-1}{1}+\binom{n-1}{2}\right]x^{n-2}y^2+\\
&\qquad\cdots+\left[\binom{n-1}{n-2}+\binom{n-1}{n-1}\right]xy^{n-1}+\binom{n-1}{n-1}y^n\\
&= x^n+\binom{n}{1}x^{n-1}y+\binom{n}{2}x^{n-2}y^2+\cdots+\binom{n}{n}y^n.
\end{aligned}$$

Chapter 10

10.1) Nontrivial ideals: $\{0,b\}$, $\{0,2a\}$, and $\{0,b,2a,2a+b\}$. Additional nontrivial subrings: $\{0,2a+b\}$, $\{0,a,2a,3a\}$, and $\{0,a+b,2a,3a+b\}$.
10.3) $R_1 : a^2 = a\cdot b = b\cdot a = b^2 = 0$; $R_2 : a^2 = b, b^2 = a\cdot b = b\cdot a = 0$; $R_3 : a^2 = a, b^2 = a\cdot b = b\cdot a = 0$; $R_4 : a^2 = a, a\cdot b = b, b^2 = b\cdot a = 0$; $R_5 : a^2 = a, b\cdot a = b, b^2 = a\cdot b = 0$; $R_6 : a^2 = b^2 = a, a\cdot b = b\cdot a = b$; $R_7 : a^2 = a, b^2 = a\cdot b = b\cdot a = b$; $R_8 : a^2 = a, a\cdot b = b\cdot a = b, b^2 = a+b$.
10.5) If $a, b \in A$, then $a\cdot y = b\cdot y = 0$, so $(a-b)\cdot y = 0$ and $(a\cdot b)\cdot y = 0$, so $a-b$ and $a\cdot b$ are in A.

10.7) If $a \in X + Y$ and $z \in R$, then $a = x + y$ for some $x \in X$ and $y \in Y$. Then $a \cdot z = (x \cdot z) + (y \cdot z) \in X + Y$. Likewise, $z \cdot a \in X + Y$.

10.9) If $a \in X \cdot Y$, and $z \in R$, then $a = x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n$, so $a \cdot z = x_1 \cdot (y_1 \cdot z) + x_2 \cdot (y_2 \cdot z) + \cdots x_n \cdot (y_n \cdot z) \in X \cdot Y$. Likewise, $z \cdot a \in X \cdot Y$.

10.11) If $a \in X \cdot Y$, then $a = x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n \in X$. Likewise, $a \in Y$, so $a \in X \cap Y$.

10.13) $\{0\}$, $\{0, a, 2a, 3a\}$, $\{0, 2a\}$, $\{0, b\}$, $\{0, a + b, 2a3a + b\}$, $\{0, 2a + b, b, 2a\}$, and the whole ring.

10.15) $\{0\}$, $\{0, a\}$, $\{0, b\}$, $\{0, c\}$, and the whole ring.

10.17) $\{0\}$, $\{0, e\}$, $\{0, a\}$, $\{0, b\}$, $\{0, c\}$, $\{0, d\}$, $\{0, f\}$, $\{0, e, c, g\}$, $\{0, e, a, d\}$, $\{0, e, b, f\}$, $\{0, a, b, c\}$, $\{0, c, d, f\}$, and the whole ring.

10.19)

$+$	$\{0, c\}$	$\{a, b\}$
$\{0, c\}$	$\{0, c\}$	$\{a, b\}$
$\{a, b\}$	$\{a, b\}$	$\{0, c\}$

$\cdot$	$\{0, c\}$	$\{a, b\}$
$\{0, c\}$	$\{0, c\}$	$\{0, c\}$
$\{a, b\}$	$\{0, c\}$	$\{a, b\}$

10.21)

$+$	$\{0, b\}$	$\{a, a + b\}$	$\{2a, 2a + b\}$	$\{3a, 3a + b\}$
$\{0, b\}$	$\{0, b\}$	$\{a, a + b\}$	$\{2a, 2a + b\}$	$\{3a, 3a + b\}$
$\{a, a + b\}$	$\{a, a + b\}$	$\{2a, 2a + b\}$	$\{3a, 3a + b\}$	$\{0, b\}$
$\{2a, 2a + b\}$	$\{2a, 2a + b\}$	$\{3a, 3a + b\}$	$\{0, b\}$	$\{a, a + b\}$
$\{3a, 3a + b\}$	$\{3a, 3a + b\}$	$\{0, b\}$	$\{a, a + b\}$	$\{2a, 2a + b\}$

$\cdot$	$\{0, b\}$	$\{a, a + b\}$	$\{2a, 2a + b\}$	$\{3a, 3a + b\}$
$\{0, b\}$	$\{0, b\}$	$\{0, b\}$	$\{0, b\}$	$\{0, b\}$
$\{a, a + b\}$	$\{0, b\}$	$\{a, a + b\}$	$\{2a, 2a + b\}$	$\{3a, 3a + b\}$
$\{2a, 2a + b\}$	$\{0, b\}$	$\{2a, 2a + b\}$	$\{0, b\}$	$\{2a, 2a + b\}$
$\{3a, 3a + b\}$	$\{0, b\}$	$\{3a, 3a + b\}$	$\{2a, 2a + b\}$	$\{a, a + b\}$

10.23) $\{0, a, b, c\}$ gives a copy of T_4 inside of T_8.

10.25) $a + b$ and $3a + b$.

10.27) Neither T_4 nor T_8 have irreducible elements.

10.29) a and $3a$ are prime, but not irreducible.

10.31) T_4^t has an element c for which $c \cdot x = 0$ for all x, T_4 has no such element.

10.33) Since a non-commutative ring must have a non-cyclic additive group, the smallest such ring would have additive group of $Z_2 \times Z_2$. If $x^2 = y$ for two nonzero elements x and y, then $x \cdot y = y \cdot x$, and the whole ring would commute. Thus, $x^2 = 0$ or x for all $x \in R$. If two nonzero elements have $x^2 = y^2 = 0$, then $x \cdot y \neq x$ or else $(x \cdot y) \cdot y = x \neq x \cdot (y \cdot y)$, likewise $x \cdot y \neq y$. Also $x \cdot y \neq x + y$, or else $x \cdot (x \cdot (x + y)) = x + y \neq (x \cdot x) \cdot (x + y)$. This means that $x \cdot y = 0$, and similarly $y \cdot x = 0$, and the ring would commute. So there are at least two elements for which $x^2 \neq 0$, call them a and b. Then $a^2 = a$, $b^2 = b$. If $(a + b)^2 = a + b$, then $a \cdot b = b \cdot a$, so we need $(a + b)^2 = 0$. Then $a \cdot b \neq a + b$, or else $(a \cdot b) \cdot b = (a + b) \cdot b = a \neq a \cdot (b \cdot b)$. Likewise, $a \cdot b \neq 0$, otherwise $(a + b)^2 = 0$ would force $b \cdot a = a + b$. So for $a \cdot b \neq b \cdot a$, one must be a, and the other b, yielding T_4 and T_4^t respectively.

10.35)

$+$	A	$1+A$	$2+A$	$3+A$	$4+A$	$5+A$
A	A	$1+A$	$2+A$	$3+A$	$4+A$	$5+A$
$1+A$	$1+A$	$2+A$	$3+A$	$4+A$	$5+A$	A
$2+A$	$2+A$	$3+A$	$4+A$	$5+A$	A	$1+A$
$3+A$	$3+A$	$4+A$	$5+A$	A	$1+A$	$2+A$
$4+A$	$4+A$	$5+A$	A	$1+A$	$2+A$	$3+A$
$5+A$	$5+A$	A	$1+A$	$2+A$	$3+A$	$4+A$

$\cdot$	A	$1+A$	$2+A$	$3+A$	$4+A$	$5+A$
A	A	A	A	A	A	A
$1+A$	A	$1+A$	$2+A$	$3+A$	$4+A$	$5+A$
$2+A$	A	$2+A$	$4+A$	A	$2+A$	$4+A$
$3+A$	A	$3+A$	A	$3+A$	A	$3+A$
$4+A$	A	$4+A$	$2+A$	A	$4+A$	$2+A$
$5+A$	A	$5+A$	$4+A$	$3+A$	$2+A$	$1+A$

10.37) $2+(8)$ is a generator of the additive group of $(2)/(8)$, but for every element x of $(2)/(8)$, $x^2=(8)$ or $4+(8)$, so there is no multiplicative identity.
10.39) $2=\phi(1\cdot 1)\neq\phi(1)\cdot\phi(1)=4$.
10.41) $\phi(x)+\phi(y)=a+c-(b+d)i=\phi(x+y)$, $\phi(x)\cdot\phi(y)=(a-bi)(c-di)=ac-bd-(bc+ad)i=\phi(x\cdot y)$.
10.43) If $a,b\in I$, then there are $x,z\in R$ such that $a=x\cdot y$ and $b=x\cdot z$. Then $a-b=(x-z)\cdot y\in I$, and if $c\in R$, then $a\cdot c=c\cdot a=(c\cdot x)\cdot y\in I$.
10.45) If a and b are nilpotent, then $a^m=b^n=0$ for some m and n. By problem 9.33, $(a-b)^{m+n}=a^{m+n}-\binom{m+n}{1}a^{m+n-1}b+\binom{m+n}{2}a^{m+n-2}b^2-\cdots+(-1)^m\binom{m+n}{m}a^mb^n+\cdots b^{m+n}=0$. So $a-b$ is nilpotient, and if $x\in R$, $(a\cdot x)^m=a^m\cdot x^m=0$, so $a\cdot x$ is nilpotient.
10.47) The homomorphism $\phi:R\mapsto R/I$, given by $\phi(x)=x+I$, restricted to the ideal K, produces $\phi':K\mapsto(K+I)/I$. The kernel of ϕ' is $K\cap I$, and so by the first isomorphism theorem for rings (10.2), $K/(K\cap I)\approx(K+I)/I$.

Chapter 11

11.1) All factorizations reveal triple roots. Reason: For real numbers, $(x+y)^3=x^3+3x^2y+3xy^2+y^3$, but since we are working mod 3, $(x+y)^3=x^3+y^3$.
11.3)

```
InitDomain[0]
Homomorph[F]
Define[F[Sqrt[5]],-Sqrt[5]]
CheckHomo[F,{1, Sqrt[5]}]
```

11.5) $(i+2)(i+3)=0$ is this ring, so it is not a field. Primes that are one more than a multiple of 4 will fail to form a field, but primes that are one less than a multiple of 4 will form a field.
11.7) $(\sqrt{17}-1)/4=\cos(2\pi/17)+\cos(4\pi/17)+\cos(8\pi/17)+\cos(16\pi/17)$.
11.9) 2.

11.11) Let the identity e have order n in the additive group. Then the characteristic cannot be less than n, but $nx = n(x \cdot e) = (ne) \cdot x = 0$ for all $x \in R$.

11.13) Let n be the order of the ring R. Then $nx = 0$ for all $x \in R$, so the characteristic would be at most n.

11.15) $\left(\frac{u}{v}\right) \cdot \left(\left(\frac{x}{y}\right) + \left(\frac{z}{w}\right)\right) = \left(\frac{u}{v}\right) \cdot \left(\frac{xw+yz}{yw}\right) = \left(\frac{uxw+uyz}{vyw}\right)$, whereas

$\left(\frac{u}{v}\right) \cdot \left(\frac{x}{y}\right) + \left(\frac{u}{v}\right) \cdot \left(\frac{z}{w}\right) = \left(\frac{ux}{vy}\right) + \left(\frac{uz}{vw}\right) = \left(\frac{uxvw+vyuz}{v^2yw}\right) = \left(\frac{uxw+uyz}{vyw}\right)$.

11.17) $(0,1) \equiv (0,2)$, $(1,1) \equiv (2,2)$, $(1,2) \equiv (2,1)$.

$+$	$\{(0,1),(0,2)\}$	$\{(1,1),(2,2)\}$	$\{(2,1),(1,2)\}$
$\{(0,1),(0,2)\}$	$\{(0,1),(0,2)\}$	$\{(1,1),(2,2)\}$	$\{(2,1),(1,2)\}$
$\{(1,1),(2,2)\}$	$\{(1,1),(2,2)\}$	$\{(2,1),(1,2)\}$	$\{(0,1),(0,2)\}$
$\{(2,1),(1,2)\}$	$\{(2,1),(1,2)\}$	$\{(0,1),(0,2)\}$	$\{(1,1),(2,2)\}$

$\cdot$	$\{(0,1),(0,2)\}$	$\{(1,1),(2,2)\}$	$\{(2,1),(1,2)\}$
$\{(0,1),(0,2)\}$	$\{(0,1),(0,2)\}$	$\{(0,1),(0,2)\}$	$\{(0,1),(0,2)\}$
$\{(1,1),(2,2)\}$	$\{(0,1),(0,2)\}$	$\{(1,1),(2,2)\}$	$\{(2,1),(1,2)\}$
$\{(2,1),(1,2)\}$	$\{(0,1),(0,2)\}$	$\{(2,1),(1,2)\}$	$\{(1,1),(2,2)\}$

11.19) x^2, x^2+1, x^2+2, x^2+x, x^2+x+1, x^2+x+2, x^2+2x, x^2+2x+1, x^2+2x+2, $2x^2$, $2x^2+1$, $2x^2+2$, $2x^2+x$, $2x^2+x+1$, $2x^2+x+2$, $2x^2+2x$, $2x^2+2x+1$, $2x^2+2x+2$.

11.21)

$$\begin{aligned} e^i &= 1 + \frac{i}{1!} + \frac{-1}{2!} + \frac{-i}{3!} + \frac{1}{4!} + \frac{i}{5!} + \cdots \\ &= \left(1 - \frac{1}{2!} + \frac{1}{4!} - \cdots\right) + i\left(\frac{1}{1!} - \frac{1}{3!} + \frac{1}{5!} - \cdots\right) = \cos 1 + i \sin 1. \end{aligned}$$

11.23)

$$1 + \frac{i}{n} = \sqrt{1 + \frac{1}{n^2}}\left(\cos(\tan^{-1}(1/n)) + i\sin(\tan^{-1}(1/n))\right),$$

so

$$\left(1 + \frac{i}{n}\right)^n = \left(1 + \frac{1}{n^2}\right)^{n/2}\left(\cos(n\tan^{-1}(1/n)) + i\sin(n\tan^{-1}(1/n))\right).$$

But

$$\lim_{n\to\infty}\left(1 + \frac{1}{n^2}\right)^{n/2} = 1 \quad \text{and} \quad \lim_{n\to\infty} n\tan^{-1}(1/n) = 1$$

by L'Hôpital's rule.

11.25) $\ln 2 - \pi/6 + 2k\pi i$, where $k \in \mathbb{Z}$.

11.27) $\sqrt{2}/2 \pm i\sqrt{2}/2$, $-\sqrt{2}/2 \pm i\sqrt{2}/2$.

11.29) $-2i$, $\pm\sqrt{3} + i$.

11.31) $\ldots$, $e^{-7\pi/4}$, $e^{-3\pi/4}$, $e^{\pi/4}$, $e^{5\pi/4}$, $e^{9\pi/4}$, $\ldots$.

11.33) $(1)^{i\ln 2/(2\pi)}$.

11.35) From DeMoivre's theorem, all solutions $z^n = 1$ are of the form $z = \cos(2k\pi/n) + i\sin(2k\pi/n) = (\cos(2\pi/n) + i\sin(2\pi/n))^k$. Thus, e_n generates the group. A generator of this group would be e_n^k, where k is coprime to n, hence a primitive n-th root of unity.
11.37) False: $(2^2)^{1/2} = 4^{1/2} = \pm 2$, yet $2^{(2\cdot 1/2)} = 2^1 = 2$.
11.39) Since $x^2 \geq 0$ and $e > 0$, then $x^2 + e > 0$.
11.41) Since $(x-y)^2 \geq 0$, $x^2 - 2xy + y^2 \geq 0$, so $x^2 + y^2 \geq 2xy$.
11.43) Let $f(x) \in \mathbb{Z}[x]^+$ have leading term ax^m and $g(x) \in \mathbb{Z}[x]^+$ have leading term bx^n. Then $f(x) \cdot g(x)$ has a leading term abx^{m+n} which is in $\mathbb{Z}[x]^+$ since $ab > 0$. $f(x) + g(x)$ will have leading term of either ax^m, bx^n, or $(a+b)x^m$, depending on whether $m > n$, $n > m$, or $n = m$. In any case $f(x) + g(x) \in \mathbb{Z}[x]^+$. Finally, either the polynomial is 0, or the leading term is either positive or negative, so the law of trichotomy holds.
11.45) For $x > 0$ in the standard ordering, then $x = (\sqrt{x})^2 > 0$ in any ordering, so there is no nonstandard ordering of $\mathbb{R}$. Hence if there were a nontrivial automorphism ϕ, then $\phi(P) = P$. Also, $\phi(1) = 1$ since the identity must map to the identity. Then $\phi(2) = \phi(1) + \phi(1) = 2$, and likewise $\phi(n) = n$ for all integers n. Then $\phi(p/q) = \phi(p)/\phi(q) = p/q$ for all rationals. If $\phi(x) = y \neq x$, then there is a rational p/q between x and y, but then $\phi(x - p/q) = y - p/q$, which contradicts $\phi(P) = P$.

Chapter 12

12.1) $f(x) = 11x^3/6 - 19x^2/2 + 50x/3 - 8$, $f(5) = 67$.
12.3) Irreducible.
12.5) In *Mathematica*:

```
InitDomain[11]
Define[a^2,6]
R = Ring[{1,a}]
CheckRing[R]
```

Since $\mathbb{Z}[\sqrt{6}]/(11)$ is a field, (11) is a prime ideal, hence 11 is prime.
12.7) $q(x) = x + 2$, $r(x) = -4x + 2$.
12.9) $f(x) = 3x^2 - 2x + 1$.
12.11) If $x^2 + 5$ has a root a in $\mathbb{R}$, then $a^2 + 5 = 0$. But $a^2 \geq 0$, so $a^2 + 5 \geq 5$. Finally, apply proposition 12.3.
12.13) $f(0) = 4(\text{Mod } 13)$, $f(1) = f(3) = f(9) = 5(\text{Mod } 13)$, $f(2) = f(5) = f(6) = 12(\text{Mod } 13)$, $f(4) = f(10) = f(12) = 3(\text{Mod } 13)$, $f(7) = f(8) = f(11) = 9(\text{Mod } 13)$, so proposition 12.3 applies.
12.15) $(x+4)(x^2+3x+3)$.
12.17) $(x+1)(x+4)(x^2+3)$.
12.19) $(2, 1+\sqrt{-5}) = \{a + b\sqrt{-5} \mid a + b = 0(\text{Mod } 2)\}$, so this is not all of $\mathbb{Z}[\sqrt{-5}]$. If $(2, 1+\sqrt{-5}) = (c)$ for some c, then c can't be a unit, but both 2 and $1+\sqrt{-5}$ must be multiples of c. This is impossible, since both 2 and $1+\sqrt{-5}$ are irreducible.
12.21) $\{0, 2, 4, 6, 8, 10\}$, $\{0, 3, 6, 9\}$.

12.23) 3 and 15 are irreducible.
12.25) 2, 3, 4, 8, 10, 14, 15, 16 are prime.
12.27) No, every nonzero would be a unit.
12.29) By letting $\mu(x)$ be the smallest n for which $x \in S_n$, then $\mu(x) \geq 0$ for all x. If $\mu(x \cdot y) = n$, then $(x \cdot y) + S_{n-1} = R$, so $(x) + S_{n-1} = R$, hence $\mu(x) \leq n = \mu(x \cdot y)$. If y is a unit, pick $q = x \cdot y^{-1}$ and $r = 0$. Otherwise, let $n = \mu(y)$, so that $x \in (y) + S_{n-1}$, that is, there is a $r \in S_{n-1}$ for which $x = y \cdot q + r$. Then $\mu(r) < n = \mu(y)$, so μ is a Euclidean valuation on R.

Now suppose R is a Euclidean domain with a valuation $\mu(x)$, and we want to show that S_n contains all nonzero elements for which $\mu(x) \leq n$. Clearly if $\mu(y) = 0$, then y is a unit, so $y \in S_0$. Suppose that it is true for all smaller values of n. If $\mu(y) = n$, then every x can be written as $y \cdot q + r$, with $\mu(r) < n$, so $r \in S_{n-1}$. Thus $R = (y) + S_{n-1}$, so $y \in S_n$. Since S_n contains all nonzero elements for which $\mu(x) \leq n$, then every element of R is in some S_n.
12.31) A PID is a UFD, so every nonzero, non-unit x can be uniquely factored into irreducible elements, so x has an irreducible factor. But in a PID, irreducible elements are prime.
12.33) In order for $f(x)$ to be a unit, it must be a constant, but since fractional constants are not allowed, the only units are ± 1. Likewise, for 2 to factor, one of the factors would be ± 1, so 2 is irreducible. But x factors as $2 \cdot x/2 = 2 \cdot 2 \cdot x/4 = \cdots$ so 2 is a factor of x an unlimited number of times.
12.35) $a = b \cdot u$ for some unit u, so $\mu(a) = \mu(b \cdot u) \geq \mu(b)$, and $\mu(b) = \mu(a \cdot u^{-1}) \geq \mu(a)$.
12.37) x^2 can be 0, 1, 2, or 4 (Mod 7), and likewise for $-6y^2$. So the sum is 0 (Mod 7) only if $x = y = 0$. Now if $(x + y\sqrt{6}) \cdot (a + b\sqrt{6})$ is a multiple of 7, then $(x^2 - 6y^2) \cdot (a^2 - 6b^2)$ is a multiple of 7, so one of these factors, say $x^2 - 6y^2$, is a multiple of 7. But then both x and y are multiples of 7, so the original factor $(x + y\sqrt{6})$ is a multiple of 7.
12.39) Since $N(a + bi) = a^2 + b^2$ is prime, proposition 12.8 shows that $a + bi$ is irreducible, hence prime.
12.41) If $a + bi$ is a factor of p, then $a - bi$ will also be a factor, so $(a + bi) \cdot (a - bi) = a^2 + b^2$ will be a factor of p. But p is prime in the ordinary sense, so $a^2 + b^2 = p$. Problem 12.40 does the other direction.
12.43) Let $\overline{q} = (1 - \sqrt{4n+1})/2$, and $\overline{x + yq} = x + y\overline{q}$. If $N(a) = \pm 1$, then $b = \overline{a}$ is such that $a \cdot b = \pm 1$, so a has an inverse. Likewise, if a has an inverse a^{-1}, then $1 = N(1) = N(a \cdot a^{-1}) = N(a) \cdot N(a^{-1})$, so $N(a) = \pm 1$. If $N(a) = p$, and $a = b \cdot c$, then $N(b) \cdot N(c) = p$, and so either b or c is a unit.
12.45) Let $t = x \cdot y^{-1} = u + vq \in \mathbb{Q}(\sqrt{5})$, and round u and v to the nearest integers i and j. If $p = i + jq$, then $N(p - t) = a^2 + ab - b^2$, where a and b are both less than $1/2$, so $|N(p-t)| \leq 3/4$. Hence $\mu(r) = |N(r)| = |N(p \cdot y - x)| = |N(p - t) \cdot N(y)| < |N(y)| = \mu(y)$.
12.47) If $a \cdot b$ is a multiple of 2, then $N(a) \cdot N(b)$ is a multiple of $N(2) = 4$, so either $N(a)$ or $N(b)$ is even, say $N(a)$. $x^2 + xy + 5y^2$ can only be even if both x and y are, so a is a multiple of 2, hence 2 is prime. To show 3 is prime,

repeat the argument, but we need to show $x^2 + xy + 5y^2$ is a multiple of 3 only if both x and y are. This can be done via a small table for $x, y \in 0, 1, 2$.

12.49) Let b be the greatest integer not exceeding $2\text{Im}(z)/\sqrt{19}$. Then $\text{Im}(z - bq) = \text{Im}(z - \sqrt{19}b/2)$ will be between 0 and $\sqrt{19}/2$. Let a be the closest integer to $\text{Re}(z - bq)$, and let $x = a + bq$. Then $0 \le \text{Im}(z - x) < \sqrt{19}/2$, and $-1/2 \le \text{Re}(z - x) \le 1/2$.

12.51) Letting $z = m^{-1}x$, we let y be as problem 12.50 so that either $|z-y| < 1$ or $|2z - y| < 1$. We can extend $N(x)$ to $\mathbb{Q}(q)$ by $N(a + bq) = a^2 + ab + 5b^2$, $a, b \in \mathbb{Q}$. In fact, $N(z) = z\overline{z} = |z|^2$. So $|m^{-1}x - y| < 1$ or $|2m^{-1}x - y| < 1$, or $|x - my| < |m|$ or $|2x - my| < |m|$. But $x - my$ and $2x - my$ are in I, and we chose m to have minimum nonzero absolute value, so either $x = my$ or $2x = my$. In the first case, we can double y to get $2x = my$.

12.53) If I is an ideal that is not a principle ideal, we can let m be the nonzero element of I with least $N(m)$, and let $x \in I$, $x \notin (m)$. From problem 12.52 we can find a y (not a multiple of 2) such that $x = (m/2)y$. Then $x\overline{y} = my\overline{y}/2 \in I$, and $y\overline{y}$ is some odd number, say $2n+1$. Since $m(2n+1)/2 = nm+m/2 \in I$, and $m \in I$, then $m/2 \in I$, but this contradicts the fact that m was chosen to have minimum $N(m)$.

Chapter 13

13.1) $\{0, 1, y^2 + y, y^2 + y + 1\}$ is a subfield of order 4, where y is the root of $x^4 + x + 1$ in the field extension. There is no subfield of order 8.

13.3) $\Phi_{15}(x) = (x^4 + x + 1)(x^4 + x^3 + 1)$. But $\Phi_5(x) = x^4 + x^3 + x^2 + x + 1$ is also irreducible.

13.5) $\Phi_8(x) = (x^2 + x + 2)(x^2 + 2x + 2)$. But $x^2 + 1$ is also irreducible.

13.7) $R = 2(C_1^3 - 3C_1C_2^2 + C_2^3 + 6C_1^2C_3 - 3C_1C_2C_3 + 9C_1C_3^2 - 3C_2C_3^2 + C_3^3 + 2C_4^3 - 6C_4C_5^2 + 2C_5^2 + 12C_4^2C_6 - 6C_4C_5C_6 + 18C_4C_6^2 - 6C_5C_6^2 + 2C_6^3 - 6C_1C_4C_7 - 12C_3C_4C_7 + 6C_2C_5C_7 + 12C_3C_5C_7 - 12C_1C_6C_7 - 6C_2C_6C_7 - 18C_3C_6C_7 + 4C_7^3 + 6C_2C_4C_8 - 6C_3C_4C_8 + 6C_1C_5C_8 - 6C_2C_5C_8 + 12C_1C_6C_8 + 6C_3C_6C_8 - 12C_7C_8^2 + 4C_8^3 - 12C_1C_4C_9 + 12C_2C_4C_9 - 18C_3C_4C_9 - 6C_1C_5C_9 + 6C_3C_5C_9 - 18C_1C_6C_9 + 6C_2C_6C_9 - 6C_3C_6C_9 + 24C_7^2C_9 - 12C_7C_8C_9 + 36C_7C_9^2 - 12C_8C_9^2 + 4C_9^3$. Since this is real, $w^{-1} = v/R$. To show $R \neq 0$, suppose $R = 0$ for some rational C_1 through C_9. But multiplying by the common denominator, we can get an integer solution to $R = 0$, and by dividing by any common factors, we can get an integer solution for which C_1 through C_9 have no common factors. Then $C_1^3 + C_1C_2^2 + C_2^3 + C_1C_2C_3 + C_1C_3^3 + C_2C_3^2 + C_3^3 = 0(\text{Mod } 2)$. The only combination for this to be true is if C_1, C_2, and C_3 are all even. Substituting $C_1 = 2B_1$, $C_2 = 2B_2$, and $C_3 = 2B_3$ into R, and factoring out 2 reveals that $C_4^3 + C_4C_5^2 + C_5^3 + C_4C_5C_6 + C_4C_6^2 + C_5C_6^2 + C_6^3 = 0(\text{Mod } 2)$. This forces C_4, C_5, and C_6 to be even, so further replacing $C_4 = 2B_4$, $C_5 = 2B_5$, and $C_6 = 2B_6$ into R, and dividing by 2, reveals $C_7^2 + C_7C_8^2 + C_8^3 + C_7C_8C_9 + C_7C_9^2 + C_8C_9^2 + C_9^3 = 0(\text{Mod } 2)$, which once again forces C_7, C_8, and C_9 to be even. But this contradicts that C_1 through C_9 have no common factors, so $R \neq 0$.

13.9) Let y be a root of $x^2 + x + 1$ in the extension field.

$+$	0	1	y	$y+1$
0	0	1	y	$y+1$
1	1	0	$y+1$	y
y	y	$y+1$	0	1
$y+1$	$y+1$	y	1	0

$\cdot$	0	1	y	$y+1$
0	0	0	0	0
1	0	1	y	$y+1$
y	0	y	$y+1$	1
$y+1$	0	$y+1$	1	y

13.11) Let y be a root of $x^2 + x + 2$ in the extension field.

$+$	0	1	2	y	$y+1$	$y+2$	$2y$	$2y+1$	$2y+2$
0	0	1	2	y	$y+1$	$y+2$	$2y$	$2y+1$	$2y+2$
1	1	2	0	$y+1$	$y+2$	y	$2y+1$	$2y+2$	$2y$
2	2	0	1	$y+2$	y	$y+1$	$2y+2$	$2y$	$2y+1$
y	y	$y+1$	$y+2$	$2y$	$2y+1$	$2y+2$	0	1	2
$y+1$	$y+1$	$y+2$	y	$2y+1$	$2y+2$	$2y$	1	2	0
$y+2$	$y+2$	y	$y+1$	$2y+2$	$2y$	$2y+1$	1	0	1
$2y$	$2y$	$2y+1$	$2y+2$	0	1	2	y	$y+1$	$y+2$
$2y+1$	$2y+1$	$2y+2$	$2y$	1	2	0	$y+1$	$y+2$	y
$2y+2$	$2y+2$	$2y$	$2y+1$	2	0	1	$y+2$	y	$y+1$

$\cdot$	0	1	2	y	$y+1$	$y+2$	$2y$	$2y+1$	$2y+2$
0	0	0	0	0	0	0	0	0	0
1	0	1	2	y	$y+1$	$y+2$	$2y$	$2y+1$	$2y+2$
2	0	2	1	$2y$	$2y+2$	$2y+1$	y	$y+2$	$y+1$
y	0	y	$2y$	$2y+1$	1	$y+1$	$y+2$	$2y+2$	2
$y+1$	0	$y+1$	$2y+2$	1	$y+2$	$2y$	2	y	$2y+1$
$y+2$	0	$y+2$	$2y+1$	$y+1$	$2y$	2	$2y+2$	1	y
$2y$	0	$2y$	y	$y+2$	2	$2y+2$	$2y+1$	$y+1$	1
$2y+1$	0	$2y+1$	$y+2$	$2y+2$	y	1	$y+1$	2	$2y$
$2y+2$	0	$2y+2$	$y+1$	2	$2y+1$	y	1	$2y$	$y+2$

13.13) The generators are $1+i$, $1+2i$, $2+i$, $2+2i$.

13.15) The Frobenius automorphism $f : x \to x^p$ must send a generator to a generator.

13.17) We can let x be the solution given from problem 13.16. Then $(x+i)(x-i) = x^2+1$ would be a multiple of p, and clearly neither $x+i$ nor $x-i$ is a multiple of p. Therefore, p is not prime in $Z[i]$.

13.19) All subfields contain the multiplicative identity, and this element generates a subfield of order p. So this subfield is in all of the subfields of F, and since it is one of the subfields, there are no other elements in the intersection.

13.21) The subfields would have to have order p^n for some prime p. Consider the polynomial $x^{(p^n)} - x$. There are at most p^n roots, but all elements from both subfields would be roots.

13.23) If n is a multiple of d, then by corollary 13.5 $p^n - 1$ is a multiple of $p^d - 1$, and so $x^{(p^n-1)} - 1$ is divisible by $x^{(p^d-1)} - 1$, and so $x^{(p^n)} - x$ is divisible by $x^{(p^d)} - x$ in $\mathbb{Z}[x]$. Since $x^{(p^n)} - x$ factors completely in F with no double roots, so does $x^{(p^d)} - x$, and these p^d elements will form a subfield since these elements are fixed by the automorphism $x \to x^{p^d}$. Problem 13.21

gives uniqueness.
13.25) A field of order p^n can be described by $Z_p[x]/(f(x))$, where $f(x)$ is an irreducible polynomial in $Z_p[x]$ of degree n. An automorphism would be determined by where it sends one of the roots of $f(x)$, and there are n possible roots. Thus, there are at most n automorphisms, and we found n Frobenius automorphisms.
13.27) $x^2 - x + 1$.
13.29) $x^4 - x^3 + x^2 - x + 1$.
13.31) $\Phi_1(x) = x - 1$ and $\Phi_2(x) = x + 1$, so assume that it is true for previous n. Plugging in $x = 0$ into proposition 13.7 gives $0^n - 1 = -1 \cdot 1 \cdots \Phi_n(0)$, so $\Phi_n(0) = 1$.
13.33) Since $x^{(p^n)} - 1 = \Phi_1(x) \cdot \Phi_p(x) \cdot \Phi_{p^2}(x) \cdots \Phi_{p^n}(x)$, and $x^{(p^{n-1})} - 1 = \Phi_1(x) \cdot \Phi_p(x) \cdot \Phi_{p^2}(x) \cdots \Phi_{p^{n-1}}(x)$, it is clear that $\Phi_{p^n}(x) = (x^{(p^n)} - 1)/(x^{(p^{n-1})} - 1) = (Y^p - 1)/(Y - 1)$, where $Y = x^{(p^{n-1})}$. Since p is prime, this is $\Phi_p(Y) = \Phi_p(x^{(p^{n-1})})$.
13.35) Let $f(x)$ be an irreducible polynomial of degree n over Z_p, and let r be a root of $f(x)$ in $GF(p^n)$. If $r^m = 1$ for some $m < p^n - 1$, then $f(x)$ cannot be a factor of $\Phi_{(p^n-1)}(x)$, lest r be a double root of $x^{(p^n-1)} - 1$, and then would contradict lemma 13.5. However, if $r^m \neq 1$ for any $m < p^n - 1$, then $f(x)$ is a factor of $x^{(p^n-1)} - 1$, yet not a factor of any $x^m - 1$ for $m < p^n - 1$, so $f(x)$ must be a factor of $\Phi_{(p^n-1)}(x)$.

Chapter 14

14.1) $\langle 41/36, -1/18, 1/4 \rangle$.
14.3) $1/(5 + \sqrt{-3}) = 5/28 - \sqrt{-3}/28$.
14.5) Splitting field $= \mathbb{Q}(a)$, where $a^3 = -a^2 + 4a - 1$; 3-dimensional extension.
14.7) Splitting field $= \mathbb{Q}(a, b)$, where $a^5 = 2$ and $b^4 = -ab^3 - a^2b^2 - a^3b - a^4$; 20-dimensional extension.
14.9) $\{1, \sqrt{2}\}$.
14.11) $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$.
14.13) $\{1, \sqrt[3]{2}, \sqrt[3]{4}\}$.
14.15) $\{1, \sqrt{3}\}$.
14.17) $x^3 - 5$.
14.19) $x^4 + 2x^2 - 1$.
14.21) $x^8 - 4x^6 + 8x^4 - 8x^2 + 2$.
14.23) $\sqrt{\sqrt{2} - 1}$, $-\sqrt{\sqrt{2} - 1}$, $\sqrt{-\sqrt{2} - 1}$, $-\sqrt{-\sqrt{2} - 1}$.
14.25) Eight roots: $\pm\sqrt{\pm\sqrt{\pm\sqrt{2} - 1} + 1}$, where each $\pm$ can be either $+$ or $-$ independently of the other $\pm$ symbols.
14.27) $\sqrt{2} + \sqrt{5}$.
14.29) $\sqrt[3]{2} + i$.
14.31) e_{15}.
14.33) $r_1 \doteq 1.25992$, $r_2 \doteq -0.62996 + 1.09112i$, $r_3 \doteq -0.62996 - 1.09112i$. $r_2^2 \doteq -0.7937 - 1.37473i$, $r_3^2 \doteq -0.7937 + 1.37473i$, $r_2 r_3 = r_1^2 \doteq 1.5874$.
14.35) Both quadratics factor in $\mathbb{Q}(\sqrt{-3})$.

14.37) $(\sqrt[3]{2} - 4\sqrt[3]{4} - 11)/43$.
14.39) $F(u)(v)$ is the smallest subfield containing both v and $F(u)$, and $F(u)$ is the smallest subfield containing both u and F. Hence $F(u)(v) = F(u, v)$, the smallest field containing u, v, and F. By symmetry, $F(v)(u) = F(u, v)$, too.
14.41) If $a = \sqrt{m} + \sqrt{n}$, then $\sqrt{m} = (a^3 - (3m + n)a)/(2n - 2m)$, and $\sqrt{n} = (a^3 - (3n + m)a)/(2m - 2n)$. So $\mathbb{Q}(\sqrt{m}, \sqrt{n})$ is in $\mathbb{Q}(a)$, and clearly $\mathbb{Q}(a)$ is in $\mathbb{Q}(\sqrt{m}, \sqrt{n})$.
14.43) $\phi_0(x) = x$, $\phi_1(\sqrt{2}) = \sqrt{2}$, $\phi_1(\sqrt{3}) = -\sqrt{3}$, $\phi_2(\sqrt{2}) = -\sqrt{2}$, $\phi_2(\sqrt{3}) = \sqrt{3}$, $\phi_3(\sqrt{2}) = -\sqrt{2}$, $\phi_3(\sqrt{3}) = -\sqrt{3}$.

Chapter 15

15.1) $\mathrm{Gal}_{\mathbb{Q}}(K) \approx D_4$, with 8 elements.
15.3) $\mathrm{Gal}_{\mathbb{Q}}(K) \approx Z_5 \ltimes Z_4$, with 20 elements.
15.5) $\mathrm{Gal}_{\mathbb{Q}}(K) \approx Z_2 \times Z_2$, with 4 elements.
15.7) $\mathrm{Gal}_{\mathbb{Q}}(K) \approx A_5$, with 60 elements.
15.9) $\mathrm{fix}(\{\phi_0\}) = \mathbb{Q}(\sqrt{2}, \sqrt{3})$, $\mathrm{fix}(\{\phi_0, \phi_1\}) = \mathbb{Q}(\sqrt{2})$, $\mathrm{fix}(\{\phi_0, \phi_2\}) = \mathbb{Q}(\sqrt{3})$, $\mathrm{fix}(\{\phi_0, \phi_3\}) = \mathbb{Q}(\sqrt{6})$, $\mathrm{fix}(\{\phi_0, \phi_1, \phi_2, \phi_3\}) = \mathbb{Q}$.
15.11) $\phi_0(x) = x$ for all x; ϕ_1 fixes $\sqrt[3]{3}$, $r_2 \leftrightarrow r_3$; ϕ_2 fixes r_2, $\sqrt[3]{3} \leftrightarrow r_3$; ϕ_3 fixes r_3, $\sqrt[3]{3} \leftrightarrow r_2$; ϕ_4: $\sqrt[3]{3} \to r_2 \to r_3 \to \sqrt[3]{3}$; ϕ_5: $\sqrt[3]{3} \to r_3 \to r_2 \to \sqrt[3]{3}$.
15.13) $\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(\sqrt{5})$, $\mathbb{Q}(\sqrt{10})$, $\mathbb{Q}(\sqrt{2}, \sqrt{5})$.
15.15) Since $Z_7^* \approx Z_6$, we can consider $\Phi_7(x) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$.
15.17) Let ϕ be an automorphism that generates the Galois group. For an element of S_4 to have order 4, it must be a 4-cycle, so ϕ is a 4-cycle of the four roots, $\phi : r_1 \to r_2 \to r_3 \to r_4 \to r_1$. Then $\phi(k) = k$, where $k = r_1^2 r_2 + r_2^2 r_3 + r_3^2 r_4 + r_4^2 r_1$. So k is in the fixed field of ϕ, and since ϕ generates the Galois group, $k \in \mathbb{Q}$.
15.19) If the Galois group is Z_5, the roots of the polynomial can be rearranged such that $r_1^2 r_2 + r_2^2 r_3 + r_3^2 r_4 + r_4^2 r_5 + r_5^2 r_1$ is rational.
15.21) One solution: $r_1 = 1.827090915$, $r_2 = 1.338261213$, $r_3 = -0.209056927$, $r_4 = -1.956295201$, $r_1^2 r_2 + r_2^2 r_3 + r_3^2 r_4 + r_4^2 r_1 = 11$.
15.23) If a is a root, then all roots are in $\mathbb{Q}(a)$, hence $|\mathrm{Gal}_{\mathbb{Q}}(F)| \leq 4$. There is an automorphism that sends a to $a^2 - 2$, and this would send $a^2 - 2$ to $(a^2 - 2)^2 - 2$, which can't be a or else a would satisfy $x^4 - 4x^2 - x - 2 = 0$. So there is an automorphism that is not of order 2, hence $\mathrm{Gal}_{\mathbb{Q}}(F) \approx Z_4$.
15.25) The first extension is of order 5, so the Galois group must contain a 5-cycle. Also, the complex conjugate automorphism switches two roots, so is a single 2-cycle. Now any 5-cycle and 2-cycle in S_5 generate all of S_5, so the Galois group is isomorphic to S_5.
15.27) S_3.
15.29) Z_2.
15.31) Z_2.
15.33) Since ϕ fixes F, and also u, then ϕ fixes $F(u)$, and hence is in $\mathrm{Gal}_{F(u)}(E)$.
15.35) $\mathrm{Gal}_F(E)$ is a finite group, so it can only have a finite number of subgroups. Since the fundamental theorem of Galois theory shows a one-to-one

correspondence between the subgroups of $\mathrm{Gal}_F(E)$ and the subfields of E containing F, there are only a finite number of such subfields.
15.37) Z_1, Z_2, Z_3, or S_3. (Possible subgroups of S_3.)
15.39) If some polynomial $f(x)$ in $\mathbb{Q}[x]$ has Galois group G, then the splitting field of $f(x)$ can be written as $\mathbb{Q}(w)$ for some w (corollary 14.4). Then $g(x) = \mathrm{Irr}_{\mathbb{Q}}(w, x)$ will have the degree n, and will have the same splitting field. Thus, the Galois group of $g(x)$ will also be G.

Bibliography

The following list not only gives the books and articles mentioned in the text, but also additional references that may help students explore related topics.

Undergraduate textbooks on Abstract Algebra

1. J. B. Fraleigh, *A First Course in Abstract Algebra,* 7th ed., Addison Wesley, Boston (2003).

2. J. A. Gallian, *Contemporary Abstract Algebra,* 6th ed., Houghton Mifflin, Boston (2006).

3. J. Gilbert and L. Gilbert, *Elements of Modern Algebra,* 4th ed., PWS Publishing Co., Boston (1996).

4. L. J. Goldstein, *Abstract Algebra, A First Course,* Prentice-Hall, Englewood Cliffs, New Jersey (1973).

5. I. N. Herstein, *Abstract Algebra,* Macmillan Publishing Company, New York (1986).

6. T. W. Hungerford, *Abstract Algebra, An Introduction,* Saunders College Publishing, Philadelphia (1990).

7. J. J. Rotman, *A First Course in Abstract Algebra,* Prentice-Hall, Upper Saddle River, New Jersey (1996).

Graduate textbooks on Abstract Algebra

8. I. N. Herstein, *Topics in Algebra,* 2nd ed., Wiley, New York (1975).

9. J. F. Humphrey, *A Course in Group Theory,* Oxford University Press, Oxford (1996).

10. D. S. Malik, J. N. Mordeson, and M. K. Sen, *Fundamentals of Abstract Algebra,* McGraw-Hill, New York (1997).

Sources for historical information

11. D. M. Burton, *The History of Mathematics, An Introduction,* 6th ed., McGraw-Hill, Boston (2007).

12. J. H. Eves, *An Introduction to the History of Mathematics,* 6th ed., Saunders College Publishing, Fort Worth (1990).

Other sources

13. J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of Finite Groups,* Clarendon Press, Oxford (1985).

14. The GAP Group, *GAP Reference Manual,* Release 4.4.12, http://www.gap-system.org.

15. I. S. Reed and G. Solomon, "Polynomial Codes over Certain Finite Fields," *SIAM Journal of Applied Math.*, vol. 8, 1960, pp. 300-304.

16. "Reed-Solomon error correction," Wikipedia, the free encyclopedia, http://en.wikipedia.org.