PRIMES OF THE FORM $x^2 + ny^2$

PURE AND APPLIED MATHEMATICS

A Wiley Series of Texts, Monographs, and Tracts

Founded by RICHARD COURANT
Editors Emeriti: MYRON B. ALLEN III, DAVID A. COX, PETER HILTON, HARRY HOCHSTADT, PETER LAX, JOHN TOLAND

A complete list of the titles in this series appears at the end of this volume.

PRIMES OF THE FORM $x^2 + ny^2$

Fermat, Class Field Theory, and Complex Multiplication

Second Edition

DAVID A. COX
Department of Mathematics
Amherst College
Amherst, Massachusetts

WILEY

Copyright © 2013 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Cox, David A.
Primes of the form $x^2 + ny^2$: Fermat, class field theory, and complex multiplication /
David A. Cox. — Second edition.
pages cm
Originally published: Primes of the form $x^2 + ny^2$, 1989.
Includes bibliographical references and index.
ISBN 978-1-118-39018-4 (cloth)
1. Numbers, Prime. 2. Mathematics. I. Title. II. Title: Primes of the form $x^2 + ny^2$.
QA246.C69 2013
512.7'23—dc23 2013000406

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

CONTENTS

CHAPTER FOUR: ADDITIONAL TOPICS

PREFACE TO THE FIRST EDITION

Several years ago, while reading Weil's *Number Theory: An Approach Through History*, I noticed a conjecture of Euler concerning primes of the form $x^2 + 14y^2$. That same week I picked up Cohn's A *Classical Invitation to Algebraic Numbers and Class Fields* and saw the same example treated from the point of view of the Hilbert class field. The coincidence made it clear that something interesting was going on, and this book is my attempt to tell the story of this wonderful part of mathematics.

I am an algebraic geometer by training, and number theory has always been more of an avocation than a profession for me. This will help explain some of the curious omissions in the book. There may also be errors of history or attribution (for which I take full responsibility), and doubtless some of the proofs can be improved. Corrections and comments are welcome!

I would like to thank my colleagues in the number theory seminars of Oklahoma State University and the Five Colleges (Amherst College, Hampshire College, Mount Holyoke College, Smith College and the University of Massachusetts) for the opportunity to present material from this book in preliminary form. Special thanks go to Dan Flath and Peter Norman for their comments on earlier versions of the manuscript. I also thank the reference librarians at Amherst College and Oklahoma Slate University for their help in obtaining books through interlibrary loan.

DAVID A. COX

Amherst, Massachsusetts
August 1989

PREFACE TO THE SECOND EDITION

The philosophy of the second edition is to preserve as much of the original text as possible. The major changes are:

- A new §15 on Shimura reciprocity has been added, based on work of Peter Stevenhagen and Alice Gee [A10, A11, A23] and Bumkyo Cho [A6].
- The fifteen sections are now organized into four chapters:
  - The original §§1–13, which present a complete solution of $p = x^2 + ny^2$, now constitute Chapters One, Two and Three.
  - The new Chapter Four consists of the original §14 (on elliptic curves) and the new §15 (on Shimura reciprocity).
- An "Additional References" section has been added to supplement the original references [1]–[112]. This section is divided into five parts:
  - The first part consists of references [A1]–[A24] that are cited in the text. These references (by no means complete) provide updates to the book.
  - The remaining four parts give some references (also not complete) for further reading that are relevant to the topics covered in Chapters One, Two, Three and Four.
- The expanded Notation section now includes all notation used in the book. Specialized notation is listed according to the page where it first appears.

The other changes to the text are very minor, mostly to enhance clarity, improve formatting, and simplify some of the proofs. One exception is the addition of new exercises: at the end of §12, Exercise 12.31 shows how Ramanujan could have derived Weber's formula for $\mathfrak{f}_1(\sqrt{-14})^2$ (thanks to Heng Huat Chan), and at the end of §14, Exercise 14.24 gives an elliptic curve primality test for Mersenne numbers due to Dick Gross [A12] (thanks to Alice Silverberg).

The web site for the book includes typographical errors and a link to supplementary exercises for §§1–3 written by Jeffrey Stopple. The URL of the web site is

`http://www.cs.amherst.edu/~dac/primes.html`

I would like to thank the following people for the errors they found in the first edition and for the suggestions they made: Michael Baake, Dominique Bernardi, Jeff Beyerl, Reinier Bröker, Tony Feng, Nicholas Gavrielides, Lee Goswik, Christian Guenther, Shiv Gupta, Kazuo Hata, Yves Hellegouarach, Norm Hurt, Tim Hutchinson, Trevor Hyde, Maurice Kostas, Susumu Kuninaga, Franz Lemmermeyer, Joseph Lipman, Mario Magioladitis, David May, Stephen Mildenhall, Takashi Ono, Frans Oort, Alf van der Poorten, Jerry Shurman, Alice Silverberg, Neil Sloane, Steve Swanson, Cihangir Tezcan, Satoshi Tomabechi, Fan Xingyuan and Noriko Yui.

Please let me know if you find any errors in the new edition!

My hope is that the second edition of *Primes of the Form* $x^2 + ny^2$ will help bring this wonderful part of number theory to a new audience of students and researchers.

DAVID A. COX

Amherst, Massachsusetts
November 2012

NOTATION

The following standard notation will be used throughout the book.

$\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$	The integers, rational numbers, real numbers, and complex numbers
$\mathrm{Re}(z)$, $\mathrm{Im}(z)$	The real and imaginary parts of $z \in \mathbb{C}$
$\mathfrak{h}$	The upper half plane $\{x+iy \in \mathbb{C} : y > 0\}$
$\mathbb{F}_q$	The finite field with q elements
$\mathbb{Z}_p$	The ring of p-adic integers
$\mathbb{Z}/n\mathbb{Z}$	The ring of integers modulo n
$[a] \in A/B$	The coset of $a \in A$ in the quotient A/B
R^*	The group of units in a commutative ring R with identity
$\mathrm{GL}(2,R)$	The group of invertible matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, $a,b,c,d \in R$
$\mathrm{SL}(2,R)$	The subgroup of $\mathrm{GL}(2,R)$ of matrices with determinant 1
I	The 2×2 identity matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$
$\mathrm{Gal}(L/K)$	The Galois group of the finite extension $K \subset L$
$[L:K]$	The degree of a the finite extension $K \subset L$
$\mathcal{O}_K$	The ring of algebraic integers in a finite extension K of $\mathbb{Q}$
$\zeta_n = e^{2\pi i/n}$	The standard primitive nth root of unity
$[a,b]$	The set $\{ma+nb : m,n \in \mathbb{Z}\}$
$\gcd(a,b)$	The greatest common divisor of the integers a and b
$\phi(n)$	The Euler ϕ-function
$\log(x)$	The logarithm to the base e of $x \in \mathbb{R}$
$[x]$	The greatest integer $\leq x$ for $x \in \mathbb{R}$
$\lvert S \rvert$	The number of elements in a finite set S
$G \rtimes H$	The semidirect product, where H acts on G
$\ker(\varphi)$, $\mathrm{im}(\varphi)$	The kernel and image of a homomorphism φ
Q.E.D.	The end of a proof or the absence of a proof

Notation for Chapter One

(a/p)	The Legendre symbol	12
(a/m)	The Jacobi symbol	15
$h(D)$	The class number	27
$C(D)$	The class group	45–46
$\chi_i(a)$, $\delta(a)$, $\epsilon(a)$	The assigned characters	49
(P/Q)	The extended Jacobi symbol	66
$\mathbb{Z}[\omega]$, $\omega = e^{2\pi i/3}$	The ring for cubic reciprocity	67
$\mathbb{Z}[i]$, $i = \sqrt{-1}$	The ring of Gaussian integers	67
$N(\alpha)$	The norm of α	67
$(\alpha/\pi)_3$, $(\alpha/\pi)_4$	The cubic and biquadratic Legendre symbols	70, 74
(f, λ)	A Gaussian period	77

Notation for Chapter Two

$N(\mathfrak{a})$	The norm of an ideal	89
I_K, P_K	The groups of ideals and principal ideals of $\mathcal{O}_K$	90
$C(\mathcal{O}_K)$	The ideal class group of $\mathcal{O}_K$	90
$e_{\mathfrak{P}\|\mathfrak{p}}$, $f_{\mathfrak{P}\|\mathfrak{p}}$	The ramification and interial degrees	90
$D_{\mathfrak{P}}$, $I_{\mathfrak{P}}$	The decomposition and inertia groups	91
d_K	The discriminant of K	92
$(D/2)$	The Kronecker symbol	93
$((L/K)/\mathfrak{P})$	The Artin symbol of $\mathfrak{P} \subset \mathcal{O}_L$	95
$((L/K)/\mathfrak{p})$	The Artin symbol of $\mathfrak{p} \subset \mathcal{O}_K$ (Abelian case)	96
$((L/K)/\cdot)$	The Artin map	97
$T(\alpha)$, $N(\alpha)$	The trace and norm of α	104
$\mathcal{O}$	An order in a quadratic field	120
$f = [\mathcal{O}_K : \mathcal{O}]$	The conductor of $\mathcal{O}$	121
$I(\mathcal{O})$, $P(\mathcal{O})$	The groups of ideals and principal ideals of $\mathcal{O}$	123
$C(\mathcal{O})$	The ideal class group of $\mathcal{O}$	123
$h(\mathcal{O})$	The class number of $\mathcal{O}$	124
$C^+(\mathcal{O})$	The narrow (or strict) ideal class group	128
$C_s(\mathcal{O})$	The signed ideal class group	129
$I(\mathcal{O}, f)$, $P(\mathcal{O}, f)$	The $\mathcal{O}$-ideals and principal $\mathcal{O}$-ideals prime to f	130
$I_K(m)$	The $\mathcal{O}_K$-ideals relatively prime to m	130
$P_{K,\mathbb{Z}}(f)$	Subgroup of $I_K(f)$ satisfying $I_K(f)/P_{K,\mathbb{Z}}(f) \simeq C(\mathcal{O})$	131
$\mathfrak{m}$	A modulus in the sense of class field theory	144
$P_{K,1}(\mathfrak{m})$	An important subgroup of $I_K(\mathfrak{m})$	145
$\Phi_{\mathfrak{m}} = \Phi_{L/K,\mathfrak{m}}$	The Artin map for the modulus $\mathfrak{m}$	145
$\mathfrak{f}(L/K)$	The class field theory conductor of $K \subset L$	146–147
$(\alpha/\mathfrak{p})_n$, $(\alpha/\mathfrak{a})_n$	The nth power Legendre symbols	149
$(\alpha, \beta/\mathfrak{p})_n$	The nth power Hilbert symbol	151
$\mathcal{P}_K$	The set of prime ideals of K	152
$\delta(S)$	The Dirichlet density of $S \subset \mathcal{P}_K$	152
$S \doteq T$	$S \subset T \cup$ finite set	154
$S_{L/K}$, $\tilde{S}_{L/K}$	The primes in K splitting completely in L and variant	154–155
$N_{L/K}$	The norm map from L to K	156–157
$K_{\mathfrak{p}}$	The completion of K at $\mathfrak{p}$	156
$\mathbf{I}_K$, $\mathbf{C}_K$	The idele group and idele class group of K	156
$\Phi_{L/K}$	The idelic Artin map	156

Notation for Chapter Three

$L = [\omega_1, \omega_2]$	A lattice in $\mathbb{C}$	182
$\wp(z) = \wp(z; L)$	The Weierstrass $\wp$-function	182
$g_2(L)$, $g_3(L)$	The coefficients in the differential equation for $\wp$	182–183
$G_r(L)$	The sum $\sum_{\omega \in L - \{0\}} 1/\omega^r$	183
$\Delta(L)$	The discriminant $g_2(L)^3 - 27 g_3(L)^2$	187
e_1, e_2, e_3	The roots of $4x^3 - g_2(L)x - g_3(L)$	187
$j(L)$	The j-invariant of L	188
$j(\tau) = j([1, \tau])$	The j-function of $\tau \in \mathfrak{h}$	190
$j(\mathfrak{a})$	The j-invariant of $\mathfrak{a} \subset \mathcal{O}$	190
$g_2(\tau)$, $g_3(\tau)$	$g_2(L)$, $g_3(L)$ for the lattice $L = [1, \tau]$	200
$\Delta(\tau)$	$\Delta(L)$ for the lattice $L = [1, \tau]$	201
$q = q(\tau)$	The function $e^{2\pi i \tau}$, $\tau \in \mathfrak{h}$	204
$\Gamma_0(m)$	The group $\left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2, \mathbb{Z}) : c \equiv 0 \bmod m \right\}$	205
$C(m)$	The matrices that give the cosets of $\Gamma_0(m) \subset \mathrm{SL}(2, \mathbb{Z})$	207
$\Phi_m(X, Y)$	The modular equation	208–209
$h(z; L)$	The Weber function used to generate ray class fields	219
$\Psi(m)$	The cardinality of $C(m)$	223
$\gamma_2(\tau)$	The cube root of the j-function	226
$\gamma_3(\tau)$	The square root of $j(\tau) - 1728$	232
$\eta(\tau)$	The Dedekind η-function	233
$\mathfrak{f}(\tau)$, $\mathfrak{f}_1(\tau)$, $\mathfrak{f}_2(\tau)$	The Weber functions	233
$\sigma(z; \tau)$	The Weierstrass σ-function	234
$\Gamma_0(2)^t$	The transpose of $\Gamma_0(2)$	241
$\zeta(z)$	The Weierstrass ζ-function	252
$H_{\mathcal{O}}(X)$, $H_D(X)$	The class equation	261
$r(\mathcal{O}, m)$	$\lvert\{\alpha \in \mathcal{O} : \alpha \text{ primitive}, N(\alpha) = m\}/\mathcal{O}^*\rvert$	263
$\Phi_{m,1}(X, X)$	The product of the multiplicity one factors of $\Phi_m(X, X)$	266
$J(d_1, d_2)$, $F(n)$	Notation for the Gross–Zagier theorem	274–275

Notation for Chapter Four

E, $E(K)$	An elliptic curve and its group of points over K	284
$\mathbb{P}^2(K)$	The projective plane over the field K	284, 304
$j(E)$	The j-invariant of E	285
$\mathrm{End}_K(E)$	The endomorphism ring of E over K	287, 289
$\deg(\alpha)$	The degree of an isogeny α	288
$Frob_q$	The Frobenius endomorphism of E over $\mathbb{F}_q$	289
$\overline{E}$	The reduction of E modulo a prime	291–292
$H(D)$	The Hurwitz class number	293
$E_0(R)$	The set of points of E over a ring R	298
$\Gamma(m)$	The congruence subgroup $\{\gamma \in \mathrm{SL}(2, \mathbb{Z}) : \gamma \equiv I \bmod m\}$	309
F_m, F	The fields of modular functions of level m and of all levels	309, 311
$f^{\gamma}(\tau)$	The action of the matrix γ on $f(\tau)$	310, 312
$\widehat{\mathbb{Z}}$, $\widehat{\mathbb{Q}}$	The profinite completion of $\mathbb{Z}$ and its tensor product with $\mathbb{Q}$	311
$\mathrm{GL}(2, \mathbb{Q})^+$	The elements of $\mathrm{GL}(2, \mathbb{Q})$ with positive determinant	311
K^{ab}	The maximal Abelian extension of K	312
Φ_K	The idelic Artin map	312, 325
$g_{\tau_0}(x)$	The matrix in $\mathrm{GL}(2, \widehat{\mathbb{Q}})$ associated to $x \in \mathbf{I}_K$	312, 321

$L_{\mathcal{O}}$, $L_{\mathcal{O},m}$	The ring class field and extended ring class field	313
$r(\tau)$	The Rogers–Ramanujan continued fraction	315
$\overline{g}_{\tau_0}(u)$	The matrix in $\mathrm{GL}(2,\mathbb{Z}/m\mathbb{Z})$ associated to $u \in (\mathcal{O}/m\mathcal{O})^*$	317
$\mathcal{O}_p$	The tensor product $\mathcal{O} \otimes_{\mathbb{Z}} \mathbb{Z}_p$	319
$\widehat{\mathcal{O}}$, $\widehat{\mathcal{O}}^*$	The ring of adeles and group of ideles over $\mathcal{O}$	319
$\widehat{K}$, $\widehat{K}^*$	The ring of adeles and group of ideles over K	320
$\mathbf{I}_K^{\mathfrak{fin}}$	The group of finite ideles of K	325
$K^*\widehat{\mathcal{O}}^*$	The subgroup of $\widehat{K}^*$ that gives $L_{\mathcal{O}}$	326
$J_{\mathcal{O},m} = K^* J^{\mathfrak{l}}_{\mathcal{O},m}$	The subgroup of $\widehat{K}^*$ that gives $L_{\mathcal{O},m}$	326
$\mathbf{I}_K^{\mathfrak{m}}$	The subgroup such that $K^*\mathbf{I}_K^{\mathfrak{m}}$ gives the ray class field of $\mathfrak{m}$	331
$\mathbf{I}_K^{\mathfrak{fin},\mathfrak{m}}$	Finite version of $\mathbf{I}_K^{\mathfrak{m}}$	331

PRIMES OF THE FORM $x^2 + ny^2$

INTRODUCTION

Most first courses in number theory or abstract algebra prove a theorem of Fermat which states that for an odd prime p,

$$p = x^2 + y^2,\ x, y \in \mathbb{Z} \iff p \equiv 1 \bmod 4.$$

This is only the first of many related results that appear in Fermat's works. For example, Fermat also states that if p is an odd prime, then

$$p = x^2 + 2y^2,\ x, y \in \mathbb{Z} \iff p \equiv 1, 3 \bmod 8$$
$$p = x^2 + 3y^2,\ x, y \in \mathbb{Z} \iff p = 3 \text{ or } p \equiv 1 \bmod 3.$$

These facts are lovely in their own right, but they also make one curious to know what happens for primes of the form $x^2 + 5y^2$, $x^2 + 6y^2$, etc. This leads to the basic question of the whole book, which we formulate as follows:

Basic Question 0.1. *Given a positive integer n, which primes p can be expressed in the form*

$$p = x^2 + ny^2$$

where x and y are integers?

We will answer this question completely, and along the way we will encounter some remarkably rich areas of number theory. The first steps will be easy, involving only

Primes of the Form $x^2 + ny^2$, Second Edition. By David A. Cox
Copyright © 2013 John Wiley & Sons, Inc.

quadratic reciprocity and the elementary theory of quadratic forms in two variables over $\mathbb{Z}$. These methods work nicely in the special cases considered above by Fermat. Using genus theory and cubic and biquadratic reciprocity, we can treat some more cases, but elementary methods fail to solve the problem in general. To proceed further, we need class field theory. This provides an abstract solution to the problem, but doesn't give explicit criteria for a particular choice of n in x^2+ny^2. The final step uses modular functions and complex multiplication to show that for a given n, there is an algorithm for answering our question of when $p = x^2+ny^2$.

This book has several goals. The first, to answer the basic question, has already been stated. A second goal is to bridge the gap between elementary number theory and class field theory. Although our basic question is simple enough to be stated in any beginning course in number theory, we will see that its solution is intimately bound up with higher reciprocity laws and class field theory. A related goal is to provide a well-motivated introduction to the classical formulation of class field theory. This will be done by carefully stating the basic theorems and illustrating their power in various concrete situations.

Let us summarize the contents of the book in more detail. We begin in Chapter One with the more elementary approaches to the problem, using the works of Fermat, Euler, Lagrange, Legendre and Gauss as a guide. In §1, we will give Euler's proofs of the above theorems of Fermat for primes of the form x^2+y^2, x^2+2y^2 and x^2+3y^2, and we will see what led Euler to discover quadratic reciprocity. We will also discuss the conjectures Euler made concerning $p = x^2+ny^2$ for $n > 3$. Some of these conjectures, such as

$$p = x^2+5y^2 \iff p \equiv 1,9 \bmod 20, \tag{0.2}$$

are similar to Fermat's theorems, while others, like

$$p = x^2+27y^2 \iff \begin{cases} p \equiv 1 \bmod 3 \text{ and } 2 \text{ is a} \\ \text{cubic residue modulo } p, \end{cases}$$

are quite unexpected. For later purposes, note that this conjecture can be written in the following form:

$$p = x^2+27y^2 \iff \begin{cases} p \equiv 1 \bmod 3 \text{ and } x^3 \equiv 2 \bmod p \\ \text{has an integer solution.} \end{cases} \tag{0.3}$$

In §2, we will study Lagrange's theory of positive definite quadratic forms. After introducing the basic concepts of reduced form and class number, we will develop an elementary form of genus theory which will enable us to prove (0.2) and similar theorems. Unfortunately, for cases like (0.3), genus theory can only prove the partial result that

$$p = \left\{ \begin{array}{c} x^2+27y^2 \\ \text{or} \\ 4x^2+2xy+7y^2 \end{array} \right\} \iff p \equiv 1 \bmod 3. \tag{0.4}$$

The problem is that x^2+27y^2 and $4x^2+2xy+7y^2$ lie in the same genus and hence can't be separated by simple congruences. We will also discuss Legendre's tentative attempts at a theory of composition.

While the ideas of genus theory and composition were already present in the works of Lagrange and Legendre, the real depth of these theories wasn't revealed until Gauss came along. In §3 we will present some basic results in Gauss' *Disquisitiones Arithmeticae*, and in particular we will study the remarkable relationship between genus theory and composition. But for our purposes, the real breakthrough came when Gauss used cubic reciprocity to prove Euler's conjecture (0.3) concerning $p=x^2+27y^2$. In §4 we will give a careful statement of cubic reciprocity, and we will explain how it can be used to prove (0.3). Similarly, biquadratic reciprocity can be used to answer our question for x^2+64y^2. We will see that Gauss clearly recognized the role of higher reciprocity laws in separating forms of the same genus. This section will also begin our study of algebraic integers, for in order to state cubic and biquadratic reciprocity, we must first understand the arithmetic of the rings $\mathbb{Z}[e^{2\pi i/3}]$ and $\mathbb{Z}[i]$.

To go further requires class field theory, which is the topic of Chapter Two. We will begin in §5 with the Hilbert class field, which is the maximal unramified Abelian extension of a given number field. This will enable us to prove the following general result:

Theorem 0.5. *Let $n \equiv 1,2 \bmod 4$ be a positive squarefree integer. Then there is an irreducible polynomial $f_n(x) \in \mathbb{Z}[x]$ such that for a prime p dividing neither n nor the discriminant of $f_n(x)$,*

$$p = x^2+ny^2 \iff \begin{cases} (-n/p)=1 \text{ and } f_n(x) \equiv 0 \bmod p \\ \text{has an integer solution.} \end{cases}$$

While the statement of Theorem 0.5 is elementary, the polynomial $f_n(x)$ is quite sophisticated: it is the minimal polynomial of a primitive element of the Hilbert class field L of $K=\mathbb{Q}(\sqrt{-n})$.

As an example of this theorem, we will study the case $n=14$. We will show that the Hilbert class field of $K=\mathbb{Q}(\sqrt{-14})$ is $L=K(\alpha)$, where $\alpha=\sqrt{2\sqrt{2}-1}$. By Theorem 0.5, this will show that for an odd prime p,

$$(0.6) \qquad p = x^2+14y^2 \iff \begin{cases} (-14/p)=1 \text{ and } (x^2+1)^2 \equiv 8 \bmod p \\ \text{has an integer solution,} \end{cases}$$

which answers our basic question for x^2+14y^2. The Hilbert class field will also enable us in §6 to give new proofs of the main theorems of genus theory.

The theory sketched so far is very nice, but there are some gaps in it. The most obvious is that the above results for x^2+27y^2 and x^2+14y^2 ((0.3) and (0.6) respectively) both follow the same format, but (0.3) does *not* follow from Theorem 0.5, for $n=27$ is *not* squarefree. There should be a unified theorem that works for *all* positive n, yet the proof of Theorem 0.5 breaks down for general n because $\mathbb{Z}[\sqrt{-n}]$ is not in general the full ring of integers in $\mathbb{Q}(\sqrt{-n})$.

The goal of §§7–9 is to show that Theorem 0.5 holds for *all* positive integers n. This, in fact, is the main theorem of the whole book. In §7 we will study the rings $\mathbb{Z}[\sqrt{-n}]$ for general n, which leads to the concept of an *order* in an imaginary quadratic field. In §8 we will summarize the main theorems of class field theory and the Čebotarev Density Theorem, and in §9 we will introduce a generalization of the Hilbert class field called the ring class field, which is a certain (possibly ramified) Abelian extension of $\mathbb{Q}(\sqrt{-n})$ determined by the order $\mathbb{Z}[\sqrt{-n}]$. Then, in Theorem 9.2, we will use the Artin Reciprocity Theorem to show that Theorem 0.5 holds for *all* $n > 0$, where the polynomial $f_n(x)$ is now the minimal polynomial of a primitive element of the above ring class field. To give a concrete example of what this means, we will apply Theorem 9.2 to the case $x^2 + 27y^2$, which will give us a class field theory proof of (0.3). In §§8 and 9 we will also discuss how class field theory is related to higher reciprocity theorems.

The major drawback to the theory presented in §9 is that it is not constructive: for a given $n > 0$, we have no idea how to find the polynomial $f_n(x)$. From (0.3) and (0.6), we know $f_{27}(x)$ and $f_{14}(x)$, but the methods used in these examples hardly generalize. Chapter Three will use the theory of complex multiplication to remedy this situation. In §10 we will study elliptic functions and introduce the idea of complex multiplication, and then in §11 we will discuss modular functions for the group $\Gamma_0(m)$ and show that the j-function can be used to generate ring class fields. As an example of the wonderful formulas that can be proved, in §12 we will give Weber's computation that

$$j(\sqrt{-14}) = 2^3\left(323 + 228\sqrt{2} + \left(231 + 161\sqrt{2}\right)\sqrt{2\sqrt{2}-1}\right)^3.$$

These methods will enable us to prove the Baker–Heegner–Stark Theorem on imaginary quadratic fields of class number 1. In §13 of the book we will discuss the class equation, which is the minimal polynomial of $j(\sqrt{-n})$. We will learn how to compute the class equation, which will lead to a constructive solution of $p = x^2 + ny^2$. We will then describe some work by Deuring and by Gross and Zagier. In 1946 Deuring proved a result about the difference of singular j-invariants, which implies an especially elegant version of our main theorem, and drawing on Deuring's work, Gross and Zagier discovered yet more remarkable properties of the class equation.

The first three chapters of the book present a complete solution to the problem of when $p = x^2 + ny^2$. In Chapter Four, we pursue two additional topics, elliptic curves in §14 and Shimura reciprocity in §15, that give a more modern approach to the study of complex multiplication. We also include applications to primality testing in §14. The new §15 discusses ideles and the field of modular functions, and replaces certain pretty but ad-hoc arguments used in §12 with a more systematic treatment based on Shimura reciprocity. We also give an unexpected application to $p = x^2 + ny^2$.

Number theory is usually taught at three levels, as an undergraduate course, a beginning graduate course, or a more advanced graduate course. These levels correspond roughly to the first three chapters of the book. Chapter One requires only beginning number theory (up to quadratic reciprocity) and a semester of abstract algebra. Since the proofs of quadratic, cubic and biquadratic reciprocity are omitted,

this book would be best suited as a supplementary text in a beginning course. For Chapter Two, the reader should know Galois theory and some basic facts about algebraic number theory (these are reviewed in §5), but no previous exposure to class field theory is assumed. The theorems of class field theory are stated without proof, so that this book would be most useful as a supplement to the topics covered in a first graduate course. Chapter Three requires a knowledge of complex analysis, but otherwise it is self-contained. (Brief but complete accounts of the Weierstrass $\wp$-function and modular functions are included in §§10 and 11.) This portion of the book should be suitable for use in a graduate seminar. The same is true for Chapter Four.

There are exercises at the end of each section, many of which consist of working out the details of arguments sketched in the text. Readers learning this material for the first time should find the exercises to be useful, while more sophisticated readers may skip them without loss of continuity.

Many important (and relevant) topics are not covered in the book. An obvious omission in Chapter One concerns forms such as $x^2 - 2y^2$, which were certainly considered by Fermat and Euler. Questions of this sort lead to Pell's equation and the class field theory of real quadratic fields. We have also ignored the problem of representing arbitrary integers, not just primes, by quadratic forms, and there are interesting questions to ask about the *number* of such representations (this material is covered in Grosswald's book [47]). In Chapter Two we give a classical formulation of class field theory, with only a brief mention of adeles and ideles. A more modern treatment can be found in Neukirch [80] or Weil [104] (see also the new §15). We also do not do justice to the use of analytic methods in number theory. For a nice introduction in the case of quadratic fields, see Zagier [111]. Our treatment of elliptic curves in Chapter Four is rather incomplete. See Husemöller [58], Knapp [A14] or Silverman [93] for the basic theory, while more advanced topics are covered by Lang [73], Shimura [90] and Silverman [A21]. At a more elemenary level, there is the wonderful book [A22] by Silverman and Tate.

There are many books which touch on the number theory encountered in studying the problem of representing primes by $x^2 + ny^2$. Four books that we particularly recommend are Cohn's *A Classical Invitation to Algebraic Numbers and Class Fields* [19], Lang's *Elliptic Functions* [73], Scharlau and Opolka's *From Fermat to Minkowski* [86], and Weil's *Number Theory: An Approach Through History* [106]. These books, as well as others to be found in the References, open up an extraordinarily rich area of mathematics. The purpose of this book is to reveal some of this richness and to encourage the reader to learn more about it.

Notes on the Second Edition

The original text of the book consisted of §§1–14. For the second edition, we added the new §15 on Shimura reciprocity described above.

As a supplement to the references for the first edition, a new section *Additional References* has been added. The new references cited in the text are indicated with a leading "A" (e.g., the references Knapp [A14], Silverman [A21], and Silverman and Tate [A22] given above). This section also contains suggestions for further reading for the four chapters.

CHAPTER ONE

FROM FERMAT TO GAUSS

§1. FERMAT, EULER AND QUADRATIC RECIPROCITY

In this section we will discuss primes of the form x^2+ny^2, where n is a fixed positive integer. Our starting point will be the three theorems of Fermat for odd primes p

$$
\begin{aligned}
&p = x^2+y^2,\ x,y\in\mathbb{Z} \iff p \equiv 1 \bmod 4\\
&p = x^2+2y^2,\ x,y\in\mathbb{Z} \iff p \equiv 1 \text{ or } 3 \bmod 8\\
&p = x^2+3y^2,\ x,y\in\mathbb{Z} \iff p = 3 \text{ or } p \equiv 1 \bmod 3
\end{aligned}
\tag{1.1}
$$

mentioned in the introduction. The goals of §1 are to prove (1.1) and, more importantly, to get a sense of what's involved in studying the equation $p = x^2+ny^2$ when $n > 0$ is arbitrary. This last question was best answered by Euler, who spent 40 years proving Fermat's theorems and thinking about how they can be generalized. Our exposition will follow some of Euler's papers closely, both in the theorems proved and in the examples studied. We will see that Euler's strategy for proving (1.1) was one of the primary things that led him to discover quadratic reciprocity, and we will also discuss some of his remarkable conjectures concerning $p = x^2+ny^2$ for $n > 3$.

Primes of the Form x^2+ny^2, Second Edition. By David A. Cox

Copyright © 2013 John Wiley & Sons, Inc.

These conjectures touch on quadratic forms, composition, genus theory, cubic and biquadratic reciprocity, and will keep us busy for the rest of the chapter.

A. Fermat

Fermat's first mention of $p = x^2 + y^2$ occurs in a 1640 letter to Mersenne [35, Vol. II, p. 212], while $p = x^2 + 2y^2$ and $p = x^2 + 3y^2$ come later, first appearing in a 1654 letter to Pascal [35, Vol. II, pp. 310–314]. Although no proofs are given in these letters, Fermat states the results as theorems. Writing to Digby in 1658, he repeats these assertions in the following form:

> Every prime number which surpasses by one a multiple of four is composed of two squares. Examples are 5, 13, 17, 29, 37, 41, etc.
>
> Every prime number which surpasses by one a multiple of three is composed of a square and the triple of another square. Examples are 7, 13, 19, 31, 37, 43, etc.
>
> Every prime number which surpasses by one or three a multiple of eight is composed of a square and the double of another square. Examples are 3, 11, 17, 19, 41, 43, etc.

Fermat adds that he has solid proofs—"firmissimis demonstralibus" [35, Vol. II, pp. 402–408 (Latin), Vol. III, pp. 314–319 (French)].

The theorems (1.1) are only part of the work that Fermat did with $x^2 + ny^2$. For example, concerning $x^2 + y^2$, Fermat knew that a positive integer N is the sum of two squares if and only if the quotient of N by its largest square factor is a product of primes congruent to 1 modulo 4 [35, Vol. III, Obs. 26, pp. 256–257], and he knew the number of different ways N can be so represented [35, Vol. III, Obs. 7, pp. 243–246]. Fermat also studied forms beyond $x^2 + y^2$, $x^2 + 2y^2$ and $x^2 + 3y^2$. For example, in the 1658 letter to Digby quoted above, Fermat makes the following conjecture about $x^2 + 5y^2$, which he admits he can't prove:

> If two primes, which end in 3 or 7 and surpass by three a multiple of four, are multiplied, then their product will be composed of a square and the quintuple of another square.
>
> Examples are the numbers 3, 7, 23, 43, 47, 67, etc. Take two of them, for example 7 and 23; their product 161 is composed of a square and the quintuple of another square. Namely 81, a square, and the quintuple of 16 equal 161.

Fermat's condition on the primes is simply that they be congruent to 3 or 7 modulo 20. In §2 we will present Lagrange's proof of this conjecture, which uses ideas from genus theory and the composition of forms.

Fermat's proofs used the method of infinite descent, but that's often all he said. As an example, here is Fermat's description of his proof for $p = x^2 + y^2$ [35, Vol. II, p. 432]:

> If an arbitrarily chosen prime number, which surpasses by one a multiple of four, is not a sum of two squares, then there is a prime number of the same form, less than the given one, and then yet a third still less, etc., descending infinitely until you arrive at the number 5, which is the least of all of this nature, from which it would follow was not the sum of two squares. From this one must infer, by deduction of the impossible, that all numbers of this form are consequently composed of two squares.

This explains the philosophy of infinite descent, but doesn't tell us how to produce the required lesser prime. We have only one complete proof by Fermat. It occurs in one of his marginal notes (the area of a right triangle with integral sides cannot be an integral square [35, Vol. III, Obs. 45, pp. 271–272]—for once the margin was big enough!). The methods of this proof (see Weil [106, p. 77] or Edwards [31, pp. 10–14] for modern expositions) do not apply to our case, so that we are still in the dark. An analysis of Fermat's approach to infinite descent appears in Bussotti [A5]. Weil's book [106] makes a careful study of Fermat's letters and marginal notes, and with some hints from Euler, he reconstructs some of Fermat's proofs. Weil's arguments are quite convincing, but we won't go into them here. For the present, we prefer to leave things as Euler found them, i.e., wonderful theorems but no proofs.

B. Euler

Euler first heard of Fermat's results through his correspondence with Goldbach. In fact, Goldbach's first letter to Euler, written in December 1729, mentions Fermat's conjecture that $2^{2^n}+1$ is always prime [40, p. 10]. Shortly thereafter, Euler read some of Fermat's letters that had been printed in Wallis' *Opera* [100] (which included the one to Digby quoted above). Euler was intrigued by what he found. For example, writing to Goldbach in June 1730, Euler comments that Fermat's four-square theorem (every positive integer is a sum of four or fewer squares) is a "non inelegans theorema" [40, p. 24]. For Euler, Fermat's assertions were serious theorems deserving of proof, and finding the proofs became a life-long project. Euler's first paper on number theory, written in 1732 at age 25, disproves Fermat's claim about $2^{2^n}+1$ by showing that 641 is a factor of $2^{32}+1$ [33, Vol. II, pp. 1–5]. Euler's interest in number theory continued unabated for the next 51 years—there was a steady stream of papers introducing many of the fundamental concepts of number theory, and even after his death in 1783, his papers continued to appear until 1830 (see [33, Vol. IV–V]). Weil's book [106] gives a survey of Euler's work on number theory (other references are Burkhardt [14], Edwards [31, Chapter 2], Scharlau and Opolka [86, Chapter 3], and the introductions to Volumes II–V of Euler's collected works [33]).

We can now present Euler's proof of the first of Fermat's theorems from (1.1):

Theorem 1.2. *An odd prime p can be written as x^2+y^2 if and only if $p \equiv 1 \bmod 4$.*

Proof. If $p = x^2+y^2$, then congruences modulo 4 easily imply that $p \equiv 1 \bmod 4$. The hard work is proving the converse. We will give a modern version of Euler's proof. Given an odd prime p, there are two basic steps to be proved:

$$\begin{aligned} &\textit{Descent Step}: \text{If } p \mid x^2+y^2,\ \gcd(x,y)=1, \text{ then } p \text{ can be written} \\ &\qquad\qquad \text{as } x^2+y^2 \text{ for some possibly different } x,\ y. \\ &\textit{Reciprocity Step}: \text{If } p \equiv 1 \bmod 4, \text{ then } p \mid x^2+y^2,\ \gcd(x,y)=1. \end{aligned}$$

It will soon become clear why we use the names "Descent" and "Reciprocity."

We'll do the Descent Step first since that's what happened historically. The argument below is taken from a 1747 letter to Goldbach [40, pp. 416–419] (see also [33,

Vol. II, pp. 295–327]). We begin with the classical identity

$$(x^2+y^2)(z^2+w^2) = (xz \pm yw)^2 + (xw \mp yz)^2 \tag{1.3}$$

(see Exercise 1.1) which enables one to express composite numbers as sums of squares. The key observation is the following lemma:

Lemma 1.4. *Suppose that N is a sum of two relatively prime squares, and that $q = x^2+y^2$ is a prime divisor of N. Then N/q is also a sum of two relatively prime squares.*

Proof. Write $N = a^2+b^2$, where a and b are relatively prime. We also have $q = x^2+y^2$, and thus q divides

$$\begin{aligned} x^2N - a^2q &= x^2(a^2+b^2) - a^2(x^2+y^2) \\ &= x^2b^2 - a^2y^2 = (xb-ay)(xb+ay). \end{aligned}$$

Since q is prime, it divides one of these two factors, and changing the sign of a if necessary, we can assume that $q \mid xb-ay$. Thus $xb-ay = dq$ for some integer d.

We claim that $x \mid a+dy$. Since x and y are relatively prime, this is equivalent to $x \mid (a+dy)y$. However,

$$\begin{aligned} (a+dy)y &= ay+dy^2 = xb-dq+dy^2 \\ &= xb - d(x^2+y^2) + dy^2 = xb - dx^2, \end{aligned}$$

which is obviously divisible by x. Furthermore, if we set $a+dy = cx$, then the above equation implies that $b = dx+cy$. Thus we have

$$\begin{aligned} a &= cx-dy \\ b &= dx+cy. \end{aligned} \tag{1.5}$$

Then, using (1.3), we obtain

$$\begin{aligned} N = a^2+b^2 &= (cx-dy)^2 + (dx+cy)^2 \\ &= (x^2+y^2)(c^2+d^2) = q(c^2+d^2). \end{aligned}$$

Thus $N/q = c^2+d^2$ is a sum of squares, and (1.5) shows that c and d must be relatively prime since a and b are. This proves the lemma. Q.E.D.

To complete the proof of the Descent Step, let p be an odd prime dividing $N = a^2+b^2$, where a and b are relatively prime. If a and b are changed by multiples of p, we still have $p \mid a^2+b^2$. We may thus assume that $|a| < p/2$ and $|b| < p/2$, which in turn implies that $N < p^2/2$. The new a and b may have a greatest common divisor $d > 1$, but p doesn't divide d, so that dividing a and b by d, we may assume that $p \mid N$, $N < p^2/2$, and $N = a^2+b^2$ where $\gcd(a,b) = 1$. Then all prime divisors $q \neq p$ of N are less than p. If q were a sum of two squares, then Lemma 1.4 would show that N/q would be a multiple of p that is again a sum of two squares. If all such

q's were sums of two squares, then repeatedly applying Lemma 1.4 would imply that p itself was of the same form. So if p is not a sum of two squares, there must be a smaller prime q with the same property. Since there is nothing to prevent us from repeating this process indefinitely, we get an infinite decreasing sequence of prime numbers. This contradiction finishes the Descent Step.

This is a classical descent argument, and as Weil argues [106, pp. 68–69], it is probably similar to what Fermat did. In §2 we will take another approach to the Descent Step, using the reduction theory of positive definite quadratic forms.

The Reciprocity Step caused Euler a lot more trouble, taking him until 1749. Euler was clearly relieved when he could write to Goldbach "Now have I finally found a valid proof" [40, pp. 493–495]. The basic idea is quite simple: since $p \equiv 1 \bmod 4$, we can write $p = 4k + 1$. Then Fermat's Little Theorem implies that

$$(x^{2k} - 1)(x^{2k} + 1) \equiv x^{4k} - 1 \equiv 0 \bmod p$$

for all $x \not\equiv 0 \bmod p$. If $x^{2k} - 1 \not\equiv 0 \bmod p$ for *one* such x, then $p \mid x^{2k} + 1$, so that p divides a sum of relatively prime squares, as desired. For us, the required x is easy to find, since $x^{2k} - 1$ is a polynomial over the field $\mathbb{Z}/p\mathbb{Z}$ and hence has at most $2k < p - 1$ roots. Euler's first proof is quite different, for it uses the calculus of finite differences—see Exercise 1.2 for details. This proves Fermat's claim (1.1) for primes of the form $x^2 + y^2$. Q.E.D.

Euler used the same two-step strategy in his proofs for $x^2 + 2y^2$ and $x^2 + 3y^2$. The Descent Steps are

$$\begin{aligned} &\text{If } p \mid x^2 + 2y^2,\ \gcd(x,y) = 1, \text{ then } p \text{ is of the form } x^2 + 2y^2 \text{ for} \\ &\qquad\qquad \text{some possibly different } x,\ y \\ &\text{If } p \mid x^2 + 3y^2,\ \gcd(x,y) = 1, \text{ then } p \text{ is of the form } x^2 + 3y^2 \text{ for} \\ &\qquad\qquad \text{some possibly different } x,\ y, \end{aligned}$$

and the Reciprocity Steps are

$$\begin{aligned} &\text{If } p \equiv 1,3 \bmod 8, \text{ then } p \mid x^2 + 2y^2,\ \gcd(x,y) = 1 \\ &\text{If } p \equiv 1 \bmod 3, \text{ then } p \mid x^2 + 3y^2,\ \gcd(x,y) = 1, \end{aligned}$$

where p is always an odd prime. In each case, the Reciprocity Step was harder to prove than the Descent Step, and Euler didn't succeed in giving complete proofs of Fermat's theorems (1.1) until 1772, 40 years after he first read about them. Weil discusses the proofs for $x^2 + 2y^2$ and $x^2 + 3y^2$ in [106, pp. 178–179, 191, and 210–212], and in Exercises 1.4 and 1.5 we will present a version of Euler's argument for $x^2 + 3y^2$.

C. $p = x^2 + ny^2$ and Quadratic Reciprocity

Let's turn to the general case of $p = x^2 + ny^2$, where n is now any positive integer. To study this problem, it makes sense to start with Euler's two-step strategy. This won't

lead to a proof, but the Descent and Reciprocity Steps will both suggest some very interesting questions for us to pursue.

The Descent Step for arbitrary $n > 0$ begins with the identity

$$(x^2+ny^2)(z^2+nw^2) = (xz \pm nyw)^2 + n(xw \mp yz)^2 \tag{1.6}$$

(see Exercise 1.1), and Lemma 1.4 generalizes easily for $n > 0$ (see Exercise 1.3). Then suppose that $p \mid x^2+ny^2$. As in the proof of the Descent Step in Theorem 1.2, we can assume that $|x|, |y| \leq p/2$. For $n \leq 3$, it follows that $x^2+ny^2 < p^2$ when p is odd, and then the argument from Theorem 1.2 shows that p is of the form x^2+ny^2 (see Exercise 1.4). One might conjecture that this holds in general, i.e., that $p \mid x^2+ny^2$ always implies $p = x^2+ny^2$. Unfortunately this fails even for $n = 5$: for example, $3 \mid 21 = 1^2+5\cdot 2^2$ but $3 \neq x^2+5y^2$. Euler knew this, and most likely so did Fermat (remember his speculations about x^2+5y^2). So the question becomes: how are prime divisors of x^2+ny^2 to be represented? As we will see in §2, the proper language for this is Lagrange's theory of quadratic forms, and a complete solution to the Descent Step will follow from the properties of reduced forms.

Turning to the Reciprocity Step for $n > 0$, the general case asks for congruence conditions on a prime p which will guarantee $p \mid x^2+ny^2$. To see what kind of congruences we need, note that the conditions of (1.1) can be unified by working modulo $4n$. Thus, given $n > 0$, we're looking for a congruence of the form $p \equiv \alpha, \beta, \ldots \bmod 4n$ which implies $p \mid x^2+ny^2$, $\gcd(x,y) = 1$. To give a modern formulation of this last condition, we first define the Legendre symbol (a/p). If a is an integer and p an odd prime, then

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & p \mid a \\ 1 & p \nmid a \text{ and } a \text{ is a quadratic residue modulo } p \\ -1 & p \nmid a \text{ and } a \text{ is a quadratic nonresidue modulo } p. \end{cases}$$

We can now restate the condition for $p \mid x^2+ny^2$ as follows:

Lemma 1.7. *Let n be a nonzero integer, and let p be an odd prime not dividing n. Then*

$$p \mid x^2+ny^2,\ \gcd(x,y) = 1 \iff \left(\frac{-n}{p}\right) = 1.$$

Proof. The basic idea is that if $x^2+ny^2 \equiv 0 \bmod p$ and $\gcd(x,y) = 1$, then y must be relatively prime to p and consequently has a multiplicative inverse modulo p. The details are left to the reader (see Exercise 1.6). Q.E.D.

The arguments of the above lemma are quite elementary, but for Euler they were not so easy—he first had to realize that quadratic residues were at the heart of the matter. This took several years, and it's fun to watch his terminology evolve: in 1744, he writes "prime divisors of numbers of the form $aa - Nbb$" [33, Vol. II, p. 216]; by 1747 this changes to "residues arising from the division of squares by the prime p" [33, Vol. II, p. 313]; and by 1751 the transition is complete—Euler now uses the terms "residua" and "non-residua" freely, with the "quadratic" being understood [33, Vol. II, p. 343].

Using Lemma 1.7, the Reciprocity Step can be restated as the following question: is there a congruence $p \equiv \alpha, \beta, \ldots \bmod 4n$ which implies $(-n/p) = 1$ when p is prime? This question also makes sense when $n < 0$, and in the following discussion n will thus be allowed to be positive or negative. We will see in Corollary 1.19 that the full answer is intimately related to the law of quadratic reciprocity, and in fact the Reciprocity Step was one of the primary things that led Euler to discover quadratic reciprocity.

Euler became intensely interested in this question in the early 1740s, and he mentions numerous examples in his letters to Goldbach. In 1744 Euler collected together his examples and conjectures in the paper *Theoremata circa divisores numerorum in hac forma paa ± qbb* contentorum [33, Vol. II, pp. 194–222]. He labels his examples as "theorems," but they are really "theorems found by induction," which is eighteenth-century parlance for conjectures based on working out some particular cases. Here are of some of Euler's conjectures, stated in modern notation:

$$
\begin{aligned}
\left(\frac{-3}{p}\right) = 1 &\iff p \equiv 1,7 \bmod 12 \\
\left(\frac{-5}{p}\right) = 1 &\iff p \equiv 1,3,7,9 \bmod 20 \\
\left(\frac{-7}{p}\right) = 1 &\iff p \equiv 1,9,11,15,23,25 \bmod 28 \\
\left(\frac{3}{p}\right) = 1 &\iff p \equiv \pm 1 \bmod 12 \\
\left(\frac{5}{p}\right) = 1 &\iff p \equiv \pm 1, \pm 11 \bmod 20 \\
\left(\frac{7}{p}\right) = 1 &\iff p \equiv \pm 1, \pm 3, \pm 9 \bmod 28,
\end{aligned}
\tag{1.8}
$$

where p is an odd prime not dividing n. In looking for a unifying pattern, the bottom three look more promising because of the $\pm$'s. If we rewrite the bottom half of (1.8) using $11 \equiv -9 \bmod 20$ and $3 \equiv -25 \bmod 28$, we obtain

$$
\begin{aligned}
\left(\frac{3}{p}\right) = 1 &\iff p \equiv \pm 1 \bmod 12 \\
\left(\frac{5}{p}\right) = 1 &\iff p \equiv \pm 1, \pm 9 \bmod 20 \\
\left(\frac{7}{p}\right) = 1 &\iff p \equiv \pm 1, \pm 25, \pm 9 \bmod 28.
\end{aligned}
$$

All of the numbers that appear are odd squares!

Before getting carried away, we should note another of Euler's conjectures:

$$
\left(\frac{6}{p}\right) = 1 \iff p \equiv \pm 1, \pm 5 \bmod 24.
$$

Unfortunately, ± 5 is not a square modulo 24, and the same thing happens for $(10/p)$ and $(14/p)$. But 3, 5 and 7 are prime, while 6, 10 and 14 are composite. Thus it makes sense to make the following conjecture for the prime case:

Conjecture 1.9. *If p and q are distinct odd primes, then*

$$\left(\frac{q}{p}\right) = 1 \iff p \equiv \pm\beta^2 \bmod 4q \text{ for some odd integer } \beta.$$

The remarkable fact is that this conjecture is equivalent to the usual statement of quadratic reciprocity:

Proposition 1.10. *If p and q are distinct odd primes, then Conjecture* 1.9 *is equivalent to*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}.$$

Proof. Let $p^* = (-1)^{(p-1)/2}p$. Then the standard properties

$$\begin{aligned} \left(\frac{-1}{p}\right) &= (-1)^{(p-1)/2} \\ \left(\frac{ab}{p}\right) &= \left(\frac{a}{p}\right)\left(\frac{b}{p}\right) \end{aligned} \tag{1.11}$$

of the Legendre symbol easily imply that quadratic reciprocity is equivalent to

$$\left(\frac{p^*}{q}\right) = \left(\frac{q}{p}\right) \tag{1.12}$$

(see Exercise 1.7). Since both sides are ± 1, it follows that quadratic reciprocity can be stated as

$$\left(\frac{q}{p}\right) = 1 \iff \left(\frac{p^*}{q}\right) = 1.$$

Comparing this to Conjecture 1.9, we see that it suffices to show

$$\left(\frac{p^*}{q}\right) = 1 \iff p \equiv \pm\beta^2 \bmod 4q,\ \beta \text{ odd}. \tag{1.13}$$

The proof of (1.13) is straightforward and is left to the reader (see Exercise 1.8). Q.E.D.

With hindsight, we can see why Euler had trouble with the Reciprocity Steps for $x^2 + 2y^2$ and $x^2 + 3y^2$: he was working out special cases of quadratic reciprocity! Exercise 1.9 will discuss which special cases were involved. We will not prove quadratic reciprocity in this section, but later in §8 we will give a proof using class field theory. Proofs of a more elementary nature can be found in most number theory texts.

The discussion leading up to Conjecture 1.9 is pretty exciting, but was it what Euler did? The answer is yes and no. To explain this, we must look more closely at Euler's 1744 paper. In addition to conjectures like (1.8), the paper also contained a series of Annotations where Euler speculated on what was happening in general. For simplicity, we will concentrate on the case of (N/p),where $N > 0$. Euler notes in Annotation 13 [33, Vol. II, p. 216] that for such N's, all of the conjectures have the form

$$\left(\frac{N}{P}\right) = 1 \iff P \equiv \pm\alpha \bmod 4N$$

for certain odd values of α. Then in Annotation 16 [33, Vol. II, pp. 216–217], Euler states that "while 1 is among the values [of the α's], yet likewise any square number, which is prime to $4N$, furnishes a suitable value for α." This is close to what we want, but it doesn't say that the odd squares fill up all possible α's when N is prime. For this, we turn to Annotation 14 [33, Vol. II, p. 216], where Euler notes that the number of α's that occur is $(1/2)\phi(N)$. When N is prime, this equals $(N-1)/2$, the number of incongruent squares modulo $4N$ relatively prime to $4N$. Thus what Euler states is fully equivalent to Conjecture 1.9. In 1875, Kronecker identified these Annotations as the first complete statement of quadratic reciprocity [68, Vol. II, pp. 3–4].

The problem is that we have to read between the lines to get quadratic reciprocity. Why didn't Euler state it more explicitly? He knew that the prime case was special, for why else would he list the prime cases before the composite ones? The answer to this puzzle, as Weil points out [106, pp. 207–209], is that Euler's real goal was to characterize the α's for *all* N, not just primes. To explain this, we need to give a modern description of the $\pm\alpha$'s. The following lemma is at the heart of the matter:

Lemma 1.14. *If $D \equiv 0, 1 \bmod 4$ is a nonzero integer, then there is a unique homomorphism $\chi : (\mathbb{Z}/D\mathbb{Z})^* \to \{\pm 1\}$ such that $\chi([p]) = (D/p)$ for odd primes p not dividing D. Furthermore,*

$$\chi([-1]) = \begin{cases} 1 & \text{when } D > 0 \\ -1 & \text{when } D < 0. \end{cases}$$

Proof. The proof will make extensive use of the Jacobi symbol. Given $m > 0$ odd and relatively prime to M, recall that the Jacobi symbol (M/m) is defined to be the product

$$\left(\frac{M}{m}\right) = \prod_{i=1}^{r} \left(\frac{M}{p_i}\right)$$

where $m = p_1 \cdots p_r$ is the prime factorization of m. Note that $(M/m) = (N/m)$ when $M \equiv N \bmod m$, and there are the multiplicative identities

$$\begin{aligned} \left(\frac{MN}{m}\right) &= \left(\frac{M}{m}\right)\left(\frac{N}{m}\right) \\ \left(\frac{M}{mn}\right) &= \left(\frac{M}{m}\right)\left(\frac{M}{n}\right) \end{aligned} \tag{1.15}$$

(see Exercise 1.10). The Jacobi symbol satisfies the following version of quadratic reciprocity:

$$\begin{aligned} \left(\frac{-1}{m}\right) &= (-1)^{(m-1)/2} \\ \left(\frac{2}{m}\right) &= (-1)^{(m^2-1)/8} \\ \left(\frac{M}{m}\right) &= (-1)^{(M-1)(m-1)/4}\left(\frac{m}{M}\right) \end{aligned} \tag{1.16}$$

(see Exercise 1.10).

For this lemma, the crucial property of the Jacobi symbol is one usually not mentioned in elementary texts: if $m \equiv n \bmod D$, where m and n are odd and positive and $D \equiv 0, 1 \bmod 4$, then

$$\left(\frac{D}{m}\right) = \left(\frac{D}{n}\right). \tag{1.17}$$

The proof is quite easy when $D \equiv 1 \bmod 4$ and $D > 0$: using quadratic reciprocity (1.16), the two sides of (1.17) become

$$\begin{aligned} &(-1)^{(D-1)(m-1)/4}\left(\frac{m}{D}\right) \\ &(-1)^{(D-1)(n-1)/4}\left(\frac{n}{D}\right). \end{aligned} \tag{1.18}$$

To compare these expressions, first note that the two Jacobi symbols are equal since $m \equiv n \bmod D$. From $D \equiv 1 \bmod 4$ we see that

$$(D-1)(m-1)/4 \equiv (D-1)(n-1)/4 \equiv 0 \bmod 2$$

since m and n are odd. Thus the signs in front of (1.18) are both $+1$, and (1.17) follows. When D is even or negative, a similar argument using the supplementary laws from (1.16) shows that (1.17) still holds (see Exercise 1.11).

It follows from (1.17) that $\chi([m]) = (D/m)$ gives a well-defined homomorphism from $(\mathbb{Z}/D\mathbb{Z})^*$ to $\{\pm 1\}$ (see Exercise 1.12), and the statement concerning $\chi([-1])$ follows from the above properties of the Jacobi symbol (see Exercise 1.12). Finally, the condition that $\chi([p]) = (D/p)$ for odd primes p determines χ uniquely follows because χ is a homomorphism and every class in $(\mathbb{Z}/D\mathbb{Z})^*$ contains a positive odd number (hence a product of odd primes) by part (a) of Exercise 1.12. Q.E.D.

The above proof made heavy use of quadratic reciprocity, which is no accident: Lemma 1.14 is in fact equivalent to quadratic reciprocity and the supplementary laws (see Exercise 1.13). For us, however, the main feature of Lemma 1.14 is that it gives a complete solution of the Reciprocity Step of Euler's strategy:

Corollary 1.19. *Let n be a nonzero integer, and let $\chi : (\mathbb{Z}/4n\mathbb{Z})^* \to \{\pm 1\}$ be the homomorphism from Lemma 1.14 when $D = -4n$. If p is an odd prime not dividing n, then the following are equivalent:*

(i) $p \mid x^2 + ny^2$, $\gcd(x,y) = 1$.

(ii) $(-n/p) = 1$.

(iii) $[p] \in \ker(\chi) \subset (\mathbb{Z}/4n\mathbb{Z})^*$.

Proof. (i) and (ii) are equivalent by Lemma 1.7, and since $(-4n/p) = (-n/p)$, (ii) and (iii) are equivalent by Lemma 1.14. Q.E.D.

To see how this solves the Reciprocity Step, write $\ker(\chi) = \{[\alpha],[\beta],[\gamma],\ldots\}$ Then $[p] \in \ker(\chi)$ is equivalent to the congruence $p \equiv \alpha,\beta,\gamma,\ldots \bmod 4n$, which is exactly the kind of condition we were looking for. Actually, Lemma 1.14 allows us to refine this a bit: when $n \equiv 3 \bmod 4$, then congruence can be taken to be of the form $p \equiv \alpha,\beta,\gamma,\ldots \bmod n$ (see Exercise 1.14). We should also note that in all cases, the usual statement of quadratic reciprocity makes it easy to compute the classes in question (see Exercise 1.15 for an example).

To see how this relates to what Euler did in 1744, let N be as in our discussion of Euler's Annotations, and let $D = 4N$ in Lemma 1.14. Then $\ker(\chi)$ consists *exactly* of Euler's $\pm\alpha$'s (when $N > 0$, the lemma also implies that $-1 \in \ker(\chi)$, which explains the $\pm$ signs). The second thing to note is that when N is odd and squarefree, $K = \ker(\chi)$ is uniquely characterized by the following four properties:

(i) K is a subgroup of index 2 in $(\mathbb{Z}/4N\mathbb{Z})^*$.

(ii) $-1 \in K$ when $N > 0$ and $-1 \notin K$ when $N < 0$.

(iii) K has period N if $N \equiv 1 \bmod 4$ and period $4N$ otherwise. (Having period $P > 0$ means that if $[a],[b] \in (\mathbb{Z}/4N\mathbb{Z})^*$, $[a] \in K$ and $a \equiv b \bmod P$, then $[b] \in K$.)

(iv) K does not have any smaller period.

For a proof of this characterization, see Weil [106, pp. 287–291]. In the Annotations to his 1744 paper, Euler gives very clear statements of (i)–(iii) (see Annotations 13–16 in [33, Vol. II, pp. 216–217]), and as for (iv), he notes that N is not a period when $N \not\equiv 1 \bmod 4$, but says nothing about the possibility of smaller periods (see Annotation 20 in [33, Vol. II, p. 219]). So Euler doesn't quite give a complete characterization of $\ker(\chi)$, but he comes incredibly close. It is a tribute to Euler's insight that he could deduce this underlying structure on the basis of examples like (1.8).

D. Beyond Quadratic Reciprocity

We will next discuss some of Euler's conjectures concerning primes of the form $x^2 + ny^2$ for $n > 3$. We start with the cases $n = 5$ and 14 (taken from his 1744 paper), for each will have something unexpected to offer us.

When $n = 5$, Euler conjectured that for odd primes $p \neq 5$,

$$\begin{aligned} p = x^2 + 5y^2 &\iff p \equiv 1, 9 \bmod 20 \\ 2p = x^2 + 5y^2 &\iff p \equiv 3, 7 \bmod 20. \end{aligned} \tag{1.20}$$

Recall from (1.8) that $p \mid x^2 + 5y^2$ is equivalent to $p \equiv 1, 3, 7, 9 \bmod 20$. Hence these four congruence classes break up into two groups $\{1, 9\}$ and $\{3, 7\}$ which have quite different representability properties. This is a new phenomenon, not encountered for $x^2 + ny^2$ when $n \leq 3$. Note also that the classes 3,7 modulo 20 are the ones that entered into Fermat's speculations on $x^2 + 5y^2$, so something interesting is going on here. In §2 we will see that this is one of the examples that led Lagrange to discover genus theory.

The case $n = 14$ is yet more complicated. Here, Euler makes the following conjecture for odd primes $\neq 7$:

$$\begin{aligned} p = \left\{ \begin{array}{l} x^2 + 14y^2 \\ 2x^2 + 7y^2 \end{array} \right\} &\iff p \equiv 1, 9, 15, 23, 25, 39 \bmod 56 \\ 3p = x^2 + 14y^2 &\iff p \equiv 3, 5, 13, 19, 27, 45 \bmod 56. \end{aligned} \tag{1.21}$$

As with (1.20), the union of the two groups of congruence classes in (1.21) describes those primes for which $(-14/p) = 1$. The new puzzle here is that we don't seem to be able to separate $x^2 + 14y^2$ from $2x^2 + 7y^2$. In §2, we will see that this is not an oversight on Euler's part, for the two quadratic forms $x^2 + 14y^2$ and $2x^2 + 7y^2$ are in the same genus and hence can't be separated by congruence classes. Another puzzle is why (1.20) uses $2p$ while (1.21) uses $3p$. In §2 we will use composition to explain these facts. One could also ask what extra condition is needed to insure $p = x^2 + 14y^2$. This lies much deeper, for as we will see in §5, it involves the Hilbert class field of $\mathbb{Q}(\sqrt{-14})$.

The final examples we want to discuss come from quite a different source, the *Tractatus de numerorum doctrina capita sedecim quae supersunt,* which Euler wrote in the period 1748–1750 [33, Vol. V, pp. 182–283]. Euler intended this work to be a basic text for number theory, in the same way that his *Introductio in analysin infinitorum* [33, Vol. VIII–IX] was the first real textbook in analysis. Unfortunately, Euler never completed the *Tractatus,* and it was first published only in 1849. Weil [106, pp. 192–196] gives a description of what's in the *Tractatus* (see also [33, Vol. V, pp. XIX–XXVI]). For us, the most interesting chapters are the two that deal with cubic and biquadratic residues. Recall that a number a is a cubic (resp. biquadratic) residue modulo p if the congruence $x^3 \equiv a \bmod p$ (resp. $x^4 \equiv a \bmod p$) has an integer solution. Euler makes the following conjectures about when 2 is a cubic or biquadratic residue modulo an odd prime p:

$$p = x^2 + 27y^2 \iff \left\{ \begin{array}{l} p \equiv 1 \bmod 3 \text{ and } 2 \text{ is a} \\ \text{cubic residue modulo } p \end{array} \right. \tag{1.22}$$

$$p = x^2 + 64y^2 \iff \left\{ \begin{array}{l} p \equiv 1 \bmod 4 \text{ and } 2 \text{ is a} \\ \text{biquadratic residue modulo } p \end{array} \right. \tag{1.23}$$

(see [33, Vol. V, pp. 250 and 258]). In §4, we will see that both of these conjectures were proved by Gauss as consequences of his work on cubic and biquadratic reciprocity.

The importance of the examples (1.20)–(1.23) is hard to overestimate. Thanks to Euler's amazing ability to find patterns, we now see some of the serious problems to be tackled (in (1.20) and (1.21)), and we have our first hint of what the final solution will look like (in (1.22) and (1.23)). Much of the next three sections will be devoted to explaining and proving these conjectures. In particular, it should be clear that we need to learn a lot more about quadratic forms. Euler left us with a magnificent series of examples and conjectures, but it remained for Lagrange to develop the language which would bring the underlying structure to light.

E. Exercises

1.1. In this exercise, we prove some identities used by Euler.

(a) Prove (1.3) and its generalization (1.6).

(b) Generalize (1.6) to find an identity of the form

$$(ax^2+cy^2)(az^2+cw^2) = (?)^2 + ac(?)^2.$$

This is due to Euler [33, Vol. I, p. 424].

1.2. Let p be prime, and let $f(x)$ be a monic polynomial of degree $d < p$. This exercise will describe Euler's proof that the congruence $f(x) \not\equiv 0 \bmod p$ has a solution. Let $\Delta f(x) = f(x+1) - f(x)$ be the difference operator.

(a) For any $k \geq 1$, show that $\Delta^k f(x)$ is an integral linear combination of $f(x), f(x+1), \ldots, f(x+k)$.

(b) Show that $\Delta^d f(x) = d!$.

(c) Euler's argument is now easy to state: if $f(x) \not\equiv 0 \bmod p$ has no solutions, then $p \mid \Delta^d f(x)$ follows from (a). By (b), this is impossible.

1.3. Let n be a positive integer.

(a) Formulate and prove a version of Lemma 1.4 when a prime $q = x^2 + ny^2$ divides a number $N = a^2 + nb^2$.

(b) Show that your proof of (a) works when $n = 3$ and $q = 4$.

1.4. In this exercise, we will prove the Descent Steps for $x^2 + 2y^2$ and $x^2 + 3y^2$.

(a) If a prime p divides $x^2 + 2y^2$, $\gcd(x,y) = 1$, then adapt the argument of Theorem 1.2 to show that $p = x^2 + 2y^2$. Hint: use Exercise 1.3.

(b) Prove that if an odd prime p divides $x^2 + 3y^2$, $\gcd(x,y) = 1$, then $p = x^2 + 3y^2$. The argument is more complicated because the Descent Step fails for $p = 2$. Thus, if it fails for some odd prime p, you have to produce

an *odd* prime $q < p$ where it also fails. Hint: part (b) of Exercise 1.3 will be useful.

1.5. If $p = 3k+1$ is prime, prove that $(-3/p) = 1$. Hint:

$$\begin{aligned} 4(x^{3k}-1) &= (x^k-1)\cdot 4(x^{2k}+x^k+1) \\ &= (x^k-1)((2x^k+1)^2+3). \end{aligned}$$

Note that Exercises 1.4(b) and 1.5 prove Fermat's theorem for x^2+3y^2.

1.6. Prove Lemma 1.7.

1.7. Use the properties (1.11) of the Legendre symbol to prove that quadratic reciprocity is equivalent to (1.12).

1.8. Prove (1.13).

1.9. In this exercise we will see how the Reciprocity Steps for x^2+y^2, x^2+2y^2 and x^2+3y^2 relate to quadratic reciprocity.

(a) Use Lemma 1.7 to show that for a prime $p > 3$,

$$p \mid x^2+3y^2,\ \gcd(x,y) = 1 \iff p \equiv 1 \bmod 3$$

is equivalent to

$$\left(\frac{-3}{p}\right) = \left(\frac{p}{3}\right).$$

By (1.12), we recognize this as part of quadratic reciprocity.

(b) Use Lemma 1.7 and the bottom line of (1.11) to show that the statements

$$\begin{aligned} p \mid x^2+y^2,\ \gcd(x,y) = 1 &\iff p \equiv 1 \bmod 4 \\ p \mid x^2+2y^2,\ \gcd(x,y) = 1 &\iff p \equiv 1,3 \bmod 8 \end{aligned}$$

are equivalent to the statements

$$\begin{aligned} \left(\frac{-1}{p}\right) &= (-1)^{(p-1)/2} \\ \left(\frac{2}{p}\right) &= (-1)^{(p^2-1)/8}. \end{aligned}$$

1.10. This exercise is concerned with the properties of the Jacobi symbol (M/m) defined in the proof of Lemma 1.14.

(a) Prove that $(M/m) = (N/m)$ when $M = N \bmod m$.

(b) Prove (1.15).

(c) Prove (1.16) using quadratic reciprocity and the two supplementary laws $(-1/p) = (-1)^{(p-1)/2}$ and $(2/p) = (-1)^{(p^2-1)/8}$. Hint: if r and s are odd, show that

$$(rs-1)/2 \equiv (r-1)/2 + (s-1)/2 \bmod 2$$
$$(r^2s^2-1)/8 \equiv (r^2-1)/8 + (s^2-1)/8 \bmod 2.$$

(d) If M is a quadratic residue modulo m, show that $(M/m) = 1$. Give an example to show that the converse is not true.

1.11. Use (1.15) and (1.16) to complete the proof of (1.17) begun in the text.

1.12. This exercise is concerned with the map $\chi\colon (\mathbb{Z}/D\mathbb{Z})^* \to \{\pm 1\}$ of Lemma 1.14. When m is odd and positive, we define $\chi([m])$ to be the Jacobi symbol (D/m).

(a) Show that any class in $(\mathbb{Z}/D\mathbb{Z})^*$ may be written as $[m]$, where m is odd and positive, and then use (1.17) to show that χ is a well-defined homomorphism on $(\mathbb{Z}/D\mathbb{Z})^*$.

(b) Show that

$$\chi([-1]) = \begin{cases} 1 & \text{if } D > 0 \\ -1 & \text{if } D < 0. \end{cases}$$

(c) If $D \equiv 1 \bmod 4$, show that

$$\chi([2]) = \begin{cases} 1 & \text{if } D \equiv 1 \bmod 8 \\ -1 & \text{if } D \equiv 5 \bmod 8. \end{cases}$$

1.13. In this exercise, we will assume that Lemma 1.14 holds for all nonzero integers $D \equiv 0, 1 \bmod 4$, and we will prove quadratic reciprocity and the supplementary laws.

(a) Let p and q be distinct odd primes, and let $q^* = (-1)^{(q-1)/2}q$. By applying the lemma with $D = q^*$, show that $(q^*/\cdot)$ induces a homomorphism from $(\mathbb{Z}/q\mathbb{Z})^*$ to $\{\pm 1\}$. Since $(\cdot/q)$ can be regarded as a homomorphism between the same two groups and $(\mathbb{Z}/q\mathbb{Z})^*$ is cyclic, conclude that the two are equal.

(b) Use similar arguments to prove the supplementary laws. Hint: apply the lemma with $D = -4$ and 8 respectively.

1.14. Use Lemma 1.14 to prove that when $n \equiv 3 \bmod 4$, there are integers $\alpha, \beta, \gamma, \ldots$ such that for an odd prime p not dividing n, $p \mid x^2 + ny^2$, $\gcd(x,y) = 1$ if and only if $p \equiv \alpha, \beta, \gamma, \ldots \bmod n$.

1.15. Use quadratic reciprocity to determine those classes in $(\mathbb{Z}/84\mathbb{Z})^*$ that satisfy $(-21/p) = 1$. This tells us when $p \mid x^2 + 21y^2$, and thus solves Reciprocity Step when $n = 21$.

1.16. In the discussion following the proof of Lemma 1.14, we stated that $K = \ker(\chi)$ is characterized by the four properties (i)–(iv). When $D = 4q$, where q is an odd prime, prove that (i) and (ii) suffice to determine K uniquely.

§2. LAGRANGE, LEGENDRE AND QUADRATIC FORMS

The study of integral quadratic forms in two variables

$$f(x,y) = ax^2 + bxy + cy^2, \quad a,b,c \in \mathbb{Z}$$

began with Lagrange, who introduced the concepts of discriminant, equivalence and reduced form. When these are combined with Gauss' notion of proper equivalence, one has all of the ingredients necessary to develop the basic theory of quadratic forms. We will concentrate on the special case of positive definite forms. Here, Lagrange's theory of reduced forms is especially nice, and in particular we will get a complete solution of the Descent Step from §1. When this is combined with the solution of the Reciprocity Step given by quadratic reciprocity, we will get immediate proofs of Fermat's theorems (1.1) as well as several new results. We will then describe an elementary form of genus theory due to Lagrange, which will enable us to prove some of Euler's conjectures from §1, and we will also be able to solve our basic question of $p = x^2 + ny^2$ for quite a few n. The section will end with some historical remarks concerning Lagrange and Legendre.

A. Quadratic Forms

Our treatment of quadratic forms is taken primarily from Lagrange's "Recherches d'Arithmétique" of 1773–1775 [69, pp. 695–795] and Gauss' *Disquisitiones Arithmeticae* of 1801 [41, §§153–226]. Most of the terminology is due to Gauss, though many of the terms he introduced refer to concepts used implicitly by Lagrange (with some important exceptions).

A first definition is that a form $ax^2 + bxy + cy^2$ is *primitive* if its coefficients a,b and c are relatively prime. Note that any form is an integer multiple of a primitive form. We will deal exclusively with primitive forms.

An integer m is *represented* by a form $f(x,y)$ if the equation

$$m = f(x,y) \tag{2.1}$$

has an integer solution in x and y. If the x and y in (2.1) are relatively prime, we say that m is *properly represented* by $f(x,y)$. Note that the basic question of the book can be restated as: which primes are represented by the quadratic form $x^2 + ny^2$?

Next, we say that two forms $f(x,y)$ and $g(x,y)$ are *equivalent* if there are integers p,q,r and s such that

$$f(x,y) = g(px+qy, rx+sy) \qquad \text{and} \qquad ps - qr = \pm 1. \tag{2.2}$$

Since $\det\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) = ps - qr = \pm 1$, this means that $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right)$ is in the group of 2×2 invertible integer matrices $\mathrm{GL}(2, \mathbb{Z})$, and it follows easily that the equivalence of forms is an equivalence relation (see Exercise 2.2). An important observation is that equivalent forms represent the same numbers, and the same is true for proper representations (see Exercise 2.2). Note also that any form equivalent to a primitive form is itself primitive (see Exercise 2.2). Following Gauss, we say that an equivalence is a *proper equivalence* if $ps - qr = 1$, i.e., $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{Z})$, and it is an *improper equivalence* if $ps - qr = -1$ [41, §158]. Since $\mathrm{SL}(2,\mathbb{Z})$ is a subgroup of $\mathrm{GL}(2,\mathbb{Z})$, it follows that proper equivalence is also an equivalence relation (see Exercise 2.2).

The notion of equivalence is due to Lagrange, though he simply said that one form "can be transformed into another of the same kind" [69, p. 723]. Neither Lagrange nor Legendre made use of proper equivalence. The terms "equivalence" and "proper equivalence" are due to Gauss [41, §157], and after stating their definitions, Gauss promises that "the usefulness of these distinctions will soon be made clear" [41, §158]. In §3 we will see that he was true to his word.

As an example of these concepts, note that the forms $ax^2 + bxy + cy^2$ and $ax^2 - bxy + cy^2$ are always improperly equivalent via the substitution $(x,y) \mapsto (x,-y)$. But are they properly equivalent? This is not obvious. We will see below that the answer is sometimes yes (for $2x^2 \pm 2xy + 3y^2$) and sometimes no (for $3x^2 \pm 2xy + 5y^2$).

There is a very nice relation between proper representations and proper equivalence:

Lemma 2.3. *A form $f(x,y)$ properly represents an integer m if and only if $f(x,y)$ is properly equivalent to the form $mx^2 + Bxy + Cy^2$ for some $B, C \in \mathbb{Z}$.*

Proof. First, suppose that $f(p,q) = m$, where p and q are relatively prime. We can find integers r and s so that $ps - qr = 1$. If $f(x,y) = ax^2 + bxy + cy^2$, then

$$\begin{aligned} f(px + ry, qx + sy) &= f(p,q)x^2 + (2apr + bps + brq + 2cqs)xy + f(r,s)y^2 \\ &= mx^2 + Bxy + Cy^2 \end{aligned}$$

is of the desired form. To prove the converse, note that $mx^2 + Bxy + Cy^2$ represents m properly by taking $(x,y) = (1,0)$, and the lemma is proved. Q.E.D.

We define the *discriminant* of $ax^2 + bxy + cy^2$ to be $D = b^2 - 4ac$. To see how this definition relates to equivalence, suppose that the forms $f(x,y)$ and $g(x,y)$ have discriminants D and D' respectively, and that

$$f(x,y) = g(px + qy, rx + sy), \qquad p,q,r,s \in \mathbb{Z}.$$

Then a straightforward calculation shows that

$$D = (ps - qr)^2 D'$$

(see Exercise 2.3), so that the two forms have the same discriminant whenever $ps - qr = \pm 1$. Thus equivalent forms have the same discriminant.

The sign of the discriminant D has a strong effect on the behavior of the form. If $f(x,y) = ax^2 + bxy + cy^2$, then we have the identity

$$4af(x,y) = (2ax + by)^2 - Dy^2. \tag{2.4}$$

If $D > 0$, then $f(x,y)$ represents both positive and negative integers, and we call the form *indefinite*, while if $D < 0$, then the form represents only positive integers or only negative ones, depending on the sign of a, and $f(x,y)$ is accordingly called *positive definite* or *negative definite* (see Exercise 2.4). Note that all of these notions are invariant under equivalence.

The discriminant D influences the form in one other way: since $D = b^2 - 4ac$, we have $D \equiv b^2 \bmod 4$, and it follows that the middle coefficient b is even (resp. odd) if and only if $D \equiv 0$ (resp. 1) mod 4.

We have the following necessary and sufficient condition for a number m to be represented by a form of discriminant D:

Lemma 2.5. *Let $D \equiv 0, 1 \bmod 4$ be an integer and m be an odd integer relatively prime to D. Then m is properly represented by a primitive form of discriminant D if and only if D is a quadratic residue modulo m.*

Proof. If $f(x,y)$ properly represents m, then by Lemma 2.3, we may assume that $f(x,y) = mx^2 + bxy + cy^2$. Thus $D = b^2 - 4mc$, and $D \equiv b^2 \bmod m$ follows easily.

Conversely, suppose that $D \equiv b^2 \bmod m$. Since m is odd, we can assume that D and b have the same parity (replace b by $b + m$ if necessary), and then $D \equiv 0, 1 \bmod 4$ implies that $D \equiv b^2 \bmod 4m$. This means that $D = b^2 - 4mc$ for some c. Then $mx^2 + bxy + cy^2$ represents m properly and has discriminant D, and the coefficients are relatively prime since m is relatively prime to D. Q.E.D.

For our purposes, the most useful version of Lemma 2.5 will be the following corollary:

Corollary 2.6. *Let n be an integer and let p be an odd prime not dividing n. Then $(-n/p) = 1$ if and only if p is represented by a primitive form of discriminant $-4n$.*

Proof. This follows immediately from Lemma 2.5 since $-4n$ is a quadratic residue modulo p if and only if $(-4n/p) = (-n/p) = 1$. Q.E.D.

This corollary is relevant to the question raised in §1 when we tried to generalize the Descent Step of Euler's strategy. Recall that we asked how to represent prime divisors of $x^2 + ny^2$, $\gcd(x,y) = 1$. Note that Corollary 2.6 gives a first answer to this question, for such primes satisfy $(-n/p) = 1$, and hence are represented by forms of discriminant $-4n$. The problem is that there are too many quadratic forms of a given discriminant. For example, if the proof of Lemma 2.5 is applied to $(-3/13) = 1$, then we see that 13 is represented by the form $13x^2 + 12xy + 3y^2$ of discriminant -12. This is not very enlightening. So to improve Corollary 2.6, we need to show that every form is equivalent to an especially simple one. Lagrange's theory of reduced forms does this and a lot more.

So far, we've dealt with arbitrary quadratic forms, but from this point on, we will specialize to the positive definite case. These forms include the ones we're most interested in (namely, x^2+ny^2 for $n>0$), and their theory has a classical simplicity and elegance. In particular, there is an especially nice notion of reduced form.

A primitive positive definite form $ax^2+bxy+cy^2$ is said to be *reduced* if

$$|b| \le a \le c, \text{ and } b \ge 0 \text{ if either } |b| = a \text{ or } a = c. \tag{2.7}$$

(Note that a and c are positive since the form is positive definite.) The basic theorem is the following:

Theorem 2.8. *Every primitive positive definite form is properly equivalent to a unique reduced form.*

Proof. The first step is to show that a given form is properly equivalent to one satisfying $|b| \le a \le c$. Among all forms properly equivalent to the given one, pick $f(x,y) = ax^2+bxy+cy^2$ so that $|b|$ is as small as possible. If $a < |b|$, then

$$g(x,y) = f(x+my,y) = ax^2 + (2am+b)xy + c'y^2$$

is properly equivalent to $f(x,y)$ for any integer m. Since $a<|b|$, we can choose m so that $|2am+b| < |b|$, which contradicts our choice of $f(x,y)$. Thus $a \ge |b|$, and $c \ge |b|$ follows similarly. If $a > c$, we need to interchange the outer coefficients, which is accomplished by the proper equivalence $(x,y) \mapsto (-y,x)$. The resulting form satisfies $|b| \le a \le c$.

The next step is to show that such a form is properly equivalent to a reduced one. By definition (2.7), the form is already reduced unless $b<0$ and $a=-b$ or $a=c$. In these exceptional cases, $ax^2-bxy+cy^2$ is reduced, so that we need only show that the two forms $ax^2 \pm bxy + cy^2$ are properly equivalent. This is done as follows:

$$a=-b : (x,y) \mapsto (x+y,y) \text{ takes } ax^2-axy+cy^2 \text{ to } ax^2+axy+cy^2.$$
$$a=c \ : (x,y) \mapsto (-y,x) \text{ takes } ax^2+bxy+ay^2 \text{ to } ax^2-bxy+ay^2.$$

The final step in the proof is to show that different reduced forms cannot be properly equivalent. This is the uniqueness part of the theorem. If $f(x,y) = ax^2+bxy+cy^2$ satisfies $|b|<a<c$, then one easily shows that

$$f(x,y) \ge (a-|b|+c)\min(x^2,y^2) \tag{2.9}$$

(see Exercise 2.7). Thus $f(x,y) \ge a-|b|+c$ whenever $xy \ne 0$, and it follows that a is the smallest nonzero value of $f(x,y)$. Furthermore, if $c>a$, then c is the next smallest number represented properly by $f(x,y)$, so that in this case the outer coefficients of a reduced form give the minimum values properly represented by any equivalent form. These observations are due to Legendre [74, Vol. I, pp. 77–78].

We now prove uniqueness. For simplicity, assume that $f(x,y) = ax^2+bxy+cy^2$ is a reduced form that satisfies the strict inequalities $|b|<a<c$. Then

$$a < c < a-|b|+c, \tag{2.10}$$

and by the above considerations, these are the three smallest numbers properly represented by $f(x,y)$. Using (2.9) and (2.10), it follows that

$$\begin{aligned} f(x,y) = a,\ \gcd(x,y) = 1 &\iff (x,y) = \pm(1,0) \\ f(x,y) = c,\ \gcd(x,y) = 1 &\iff (x,y) = \pm(0,1) \end{aligned} \tag{2.11}$$

(see Exercise 2.8). Now let $g(x,y)$ be a reduced form equivalent to $f(x,y)$. Since these forms represent the same numbers and are reduced, they must have the same first coefficient a by Legendre's observation. Now consider the third coefficient c' of $g(x,y)$. We know that $a \le c'$ since $g(x,y)$ is reduced. If equality occurred, then the equation $g(x,y) = a$ would have four proper solutions $\pm(1,0)$ and $\pm(0,1)$. Since $f(x,y)$ is equivalent to $g(x,y)$, this would contradict (2.11). Thus $a < c'$, and then Legendre's observation shows that $c = c'$. Hence the outer coefficients of $f(x,y)$ and $g(x,y)$ are the same, and since they have the same discriminant, it follows that $g(x,y) = ax^2 \pm bxy + cy^2$.

It remains to show that $f(x,y) = g(x,y)$ when we make the stronger assumption that the forms are properly equivalent. If we assume that

$$g(x,y) = f(px+qy, rx+sy), \qquad ps - qr = 1,$$

then $a = g(1,0) = f(p,r)$ and $c = g(0,1) = f(q,s)$ are proper representations. By (2.11), it follows that $(p,r) = \pm(1,0)$ and $(q,s) = \pm(0,1)$. Then $ps - qr = 1$ implies $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) = \pm\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and $f(x,y) = g(x,y)$ follows easily.

When $a = |b|$ or $a = c$, the above argument breaks down, because the values in (2.10) are no longer distinct. Nevertheless, one can still show that $f(x,y)$ and $g(x,y)$ reduce to $ax^2 \pm bxy + cy^2$, and then the restriction $b \ge 0$ in definition (2.7) implies equality. (See Exercise 2.8, or for the complete details, Scharlau and Opolka [86, pp. 36–38].) Q.E.D.

Note that we can now answer our earlier question about equivalence versus proper equivalence. Namely, the forms $3x^2 \pm 2xy + 5y^2$ are clearly equivalent, but since they are both reduced, Theorem 2.8 implies that they are not properly equivalent. On the other hand, of $2x^2 \pm 2xy + 3y^2$, only $2x^2 + 2xy + 3y^2$ is reduced (because $a = |b|$), and by the proof of Theorem 2.8, it is properly equivalent to $2x^2 - 2xy + 3y^2$.

In order to complete the elementary theory of reduced forms, we need one more observation. Suppose that $ax^2 + bxy + cy^2$ is a reduced form of discriminant $D < 0$. Then $b^2 \le a^2$ and $a \le c$, so that

$$-D = 4ac - b^2 \ge 4a^2 - a^2 = 3a^2$$

and thus

$$a \le \sqrt{(-D)}/3. \tag{2.12}$$

If D is fixed, then $|b| \le a$ and (2.12) imply that there are only finitely many choices for a and b. Since $b^2 - 4ac = D$, the same is true for c, so that there are only a finite number of reduced forms of discriminant D. Then Theorem 2.8 implies that

the number of proper equivalence classes is also finite. Following Gauss [41, §223], we say that two forms are in the same *class* if they are properly equivalent. We will let $h(D)$ denote the number of classes of primitive positive definite forms of discriminant D, which by Theorem 2.8 is just the number of reduced forms. We have thus proved the following theorem:

Theorem 2.13. *Let $D < 0$ be fixed. Then the number $h(D)$ of classes of primitive positive definite forms of discriminant D is finite, and furthermore $h(D)$ is equal to the number of reduced forms of discriminant D.* Q.E.D.

The above discussion shows that there is an algorithm for computing reduced forms and class numbers which, for small discriminants, is easily implemented on a computer (see Exercise 2.9). Here are some examples that will prove useful later on:

(2.14)

D	$h(D)$	Reduced Forms of Discriminant D
-4	1	x^2+y^2
-8	1	x^2+2y^2
-12	1	x^2+3y^2
-20	2	$x^2+5y^2, 2x^2+2xy+3y^2$
-28	1	x^2+7y^2
-56	4	$x^2+14y^2, 2x^2+7y^2, 3x^2\pm 2xy+5y^2$
-108	3	$x^2+27y^2, 4x^2\pm 2xy+7y^2$
-256	4	$x^2+64y^2, 4x^2+4xy+17y^2, 5x^2\pm 2xy+13y^2$

Note, by the way, that x^2+ny^2 is always a reduced form! For a further discussion of the computational aspects of class numbers, see Buell [12] and Shanks [89] (the algorithm described in [89] makes nice use of the theory to be described in §3).

This completes our discussion of positive definite forms. We should also mention that there is a corresponding theory for indefinite forms. Its roots reach back to Fermat and Euler (both considered special cases, such as x^2-2y^2), and Lagrange and Gauss each developed a general theory of such forms. There are notions of reduced form, class number, etc., but the uniqueness problem is much more complicated. As Gauss notes, "it can happen that many reduced forms are properly equivalent among themselves" [41, §184]. Determining exactly which reduced forms are properly equivalent is not easy (see Lagrange [69, pp. 728–740] and Gauss [41, §§183–193]). There are also connections with continued fractions and Pell's equation (see [41, §§183–205]), so that the indefinite case has a very different flavor. Two modern references are Flath [36, Chapter IV] and Zagier [111, §§8, 13 and 14].

B. $p = x^2 + ny^2$ and Quadratic Forms

We can now apply the theory of positive definite quadratic forms to solve some of the problems encountered in §1. We start by giving a complete solution of the Descent Step of Euler's strategy:

Proposition 2.15. *Let n be a positive integer and p be an odd prime not dividing n. Then $(-n/p) = 1$ if and only if p is represented by one of the $h(-4n)$ reduced forms of discriminant $-4n$.*

Proof. This follows immediately from Corollary 2.6 and Theorem 2.8. Q.E.D.

In §1 we showed how quadratic reciprocity gives a general solution of the Reciprocity Step of Euler's strategy. Having just solved the Descent Step, it makes sense to put the two together and see what we get. But rather than just treat the case of forms of discriminant $-4n$, we will state a result that applies to *all* negative discriminants $D < 0$. Recall from Lemma 1.14 that there is a homomorphism $\chi : (\mathbb{Z}/D\mathbb{Z})^* \to \{\pm 1\}$ such that $\chi([p]) = (D/p)$ for odd primes not dividing D. Note that $\ker(\chi) \subset (\mathbb{Z}/D\mathbb{Z})^*$ is a subgroup of index 2. We then have the following general theorem:

Theorem 2.16. *Let $D \equiv 0, 1 \bmod 4$ be negative, and let $\chi : (\mathbb{Z}/D\mathbb{Z})^* \to \{\pm 1\}$ be the homomorphism from Lemma 1.14. Then, for an odd prime p not dividing D, $[p] \in \ker(\chi)$ if and only if p is represented by one of the $h(D)$ reduced forms of discriminant D.*

Proof. The definition of χ tells us that $[p] \in \ker(\chi)$ if and only if $(D/p) = 1$. By Lemma 2.5, this last condition is equivalent to being represented by a primitive positive definite form of discriminant D, and then we are done by Theorem 2.8. Q.E.D.

The basic content of this theorem is that there is a congruence $p \equiv \alpha, \beta, \gamma, \ldots \bmod D$ which gives necessary and sufficient conditions for an odd prime p to be represented by a reduced form of discriminant D. This result is very computational, for we know how to find the reduced forms, and quadratic reciprocity makes it easy to find the congruence classes $\alpha, \beta, \gamma, \ldots \bmod D$ such that $(D/p) = 1$.

For an example of how Theorem 2.16 works, note that $x^2 + y^2$, $x^2 + 2y^2$ and $x^2 + 3y^2$ are the only reduced forms of discriminants -4, -8 and -12 respectively (this is from (2.14)). Using quadratic reciprocity to find the congruence classes for which $(-1/p)$, $(-2/p)$ and $(-3/p)$ equal 1, we get immediate proofs of Fermat's three theorems (1.1) (see Exercise 2.11). This shows just how powerful a theory we have: Fermat's theorems are now reduced to the status of an exercise. We can also go beyond Fermat, for notice that by (2.14), $x^2 + 7y^2$ is the only reduced form of discriminant -28, and it follows easily that

$$p = x^2 + 7y^2 \iff p \equiv 1, 9, 11, 15, 23, 25 \bmod 28 \tag{2.17}$$

for primes $p \neq 7$ (see Exercise 2.11). Thus we have made significant progress in answering our basic question of when $p = x^2 + ny^2$.

Unfortunately, this method for characterizing $p = x^2 + ny^2$ works only when $h(-4n) = 1$. In 1903, Landau proved a conjecture of Gauss that there are very few n's with this property:

Theorem 2.18. *Let n be a positive integer. Then*

$$h(-4n) = 1 \iff n = 1, 2, 3, 4 \text{ or } 7.$$

Proof. We will follow Landau's proof [70]. The basic idea is very simple: $x^2 + ny^2$ is a reduced form, and for $n \notin \{1,2,3,4,7\}$, we will produce a second reduced form of the same discriminant, showing that $h(-4n) > 1$. We may assume $n > 1$.

First suppose that n is not a prime power. Then n can be written $n = ac$, where $1 < a < c$ and $\gcd(a,c) = 1$ (see Exercise 2.12), and the form

$$ax^2+cy^2$$

is reduced of disciminant $-4ac = -4n$. Thus $h(-4n) > 1$ when n is not a prime power.

Next suppose that $n = 2^r$. If $r \geq 4$, then

$$4x^2 + 4xy + (2^{r-2}+1)y^2$$

has relatively prime coefficients and is reduced since $4 \leq 2^{r-2}+1$. Furthermore, it has discriminant $4^2 - 4 \cdot 4(2^{r-2}+1) = -16 \cdot 2^{r-2} = -4n$. Thus $h(-4n) > 1$ when $n = 2^r$, $r \geq 4$. One computes directly that $h(-4 \cdot 8) = 2$ (see Exercise 2.12), which leaves us with the known cases $n = 2$ and 4.

Finally, assume that $n = p^r$, where p is an odd prime. If $n+1$ can be written $n+1 = ac$, where $2 \leq a < c$ and $\gcd(a,c) = 1$, then

$$ax^2+2xy+cy^2$$

is reduced of discriminant $2^2 - 4ac = 4 - 4(n+1) = -4n$. Thus $h(-4n) > 1$ when $n+1$ is not a prime power. But $n = p^r$ is odd, so that $n+1$ is even, and hence it remains to consider the case $n+1 = 2^s$. If $s \geq 6$, then

$$8x^2 + 6xy + (2^{s-3}+1)y^2$$

has relatively prime coefficients and is reduced since $8 \leq 2^{s-3}+1$. Furthermore, it has discriminant $6^2 - 4 \cdot 8(2^{s-3}+1) = 4 - 4 \cdot 2^s = 4 - 4(n+1) = -4n$, and hence $h(-4n) > 1$ when $s \geq 6$. The cases $s = 1,2,3,4$ and 5 correspond to $n = 1,3,7,15$ and 31 respectively. Now $n = 15$ is not a prime power, and one easily computes that $h(-4 \cdot 31) = 3$ (see Exercise 2.12). This leaves us with the three known cases $n = 1,3$ and 7, and completes the proof of the theorem. Q.E.D.

Note that we've already discussed the cases $n = 1$, 2, 3 and 7, and the case $n = 4$ was omitted since $p = x^2 + 4y^2$ is a trivial corollary of $p = x^2 + y^2$ (p is odd, so that one of x or y must be even). One could also ask if there is a similar finite list of *odd* discriminants $D < 0$ with $h(D) = 1$. The answer is yes, but the proof is *much* more difficult. We will discuss this problem in §7 and give a proof in §12.

C. Elementary Genus Theory

One consequence of Theorem 2.18 is that we need some new ideas to characterize $p = x^2 + ny^2$ when $h(-4n) > 1$. To get a sense of what's involved, consider the example $n = 5$. Here, Theorem 2.16, quadratic reciprocity and (2.14) tell us that

$$\begin{aligned} p \equiv 1,3,7,9 \bmod 20 &\iff \left(\frac{-5}{p}\right) = 1 \\ &\iff p = x^2 + 5y^2 \text{ or } 2x^2+2xy+3y^2. \end{aligned} \tag{2.19}$$

We need a method of separating reduced forms of the same discriminant, and this is where genus theory comes in. The basic idea is due to Lagrange, who, like us, used quadratic forms to prove conjectures of Fermat and Euler. But rather than working with reduced forms collectively, as we did in Theorem 2.16, Lagrange considers the congruence classes represented in $(\mathbb{Z}/D\mathbb{Z})^*$ by a single form, and he groups together forms that represent the same classes. This turns out to be the basic idea of genus theory!

Let's work out some examples to see how this grouping works. When $D = -20$, one easily computes that

$$\begin{aligned} x^2+5y^2 &\quad \text{represents} \quad 1,9 \quad \text{in} \quad (\mathbb{Z}/20\mathbb{Z})^* \\ 2x^2+2xy+3y^2 &\quad \text{represents} \quad 3,7 \quad \text{in} \quad (\mathbb{Z}/20\mathbb{Z})^* \end{aligned} \tag{2.20}$$

while for $D = -56$ one has

$$\begin{aligned} x^2+14y^2, 2x^2+7y^2 &\quad \text{represent} \quad 1,9,15.23,25,39 \quad \text{in} \quad (\mathbb{Z}/56\mathbb{Z})^* \\ 3x^2 \pm 2xy + 5y^2 &\quad \text{represent} \quad 3,5,13,19,27,45 \quad \text{in} \quad (\mathbb{Z}/56\mathbb{Z})^* \end{aligned} \tag{2.21}$$

(see Exercise 2.14—the reduced forms are taken from (2.14)). In his memoir on quadratic forms, Lagrange gives a systematic procedure for determining the congruence classes in $(\mathbb{Z}/D\mathbb{Z})^*$ represented by a form of discriminant D [69, pp. 759–765], and he includes a table listing various reduced forms together with the corresponding congruence classes [69, pp. 766–767]. The examples in Lagrange's table show that this is a very natural way to group forms of the same discriminant.

In general, we say that two primitive positive definite forms of discriminant D are in the same *genus* if they represent the same values in $(\mathbb{Z}/D\mathbb{Z})^*$. Note that equivalent forms represent the same numbers and hence are in the same genus. In particular, each genus consists of a finite number of classes of forms. The above examples show that when $D = -20$, there are two genera, each consisting of a single class, and when $D = -56$, there are again two genera, but this time each genus consists of two classes.

The real impact of this theory becomes clear when we combine it with Theorem 2.16. The basic idea is that genus theory refines our earlier correspondence between congruence classes and representations by reduced forms. For example, when $D = -20$, (2.19) tells us that $p \equiv 1,3,7,9 \bmod 20 \iff x^2 + 5y^2$ or $2x^2 + 2xy + 3y^2$.

If we combine this with (2.20), we obtain

$$\begin{aligned} p = x^2 + 5y^2 &\iff p \equiv 1,9 \bmod 20 \\ p = 2x^2 + 2xy + 3y^2 &\iff p \equiv 3,7 \bmod 20 \end{aligned} \tag{2.22}$$

when $p \neq 5$ is odd. Note that the top line of (2.22) solves Euler's conjecture (1.20) for when $p = x^2 + 5y^2$! The thing that makes this work is that the two genera represent disjoint values in $(\mathbb{Z}/20\mathbb{Z})^*$. Looking at (2.21), we see that the same thing happens when $D = -56$, and then using Theorem 2.16 it is straightforward to prove that

$$\begin{aligned} p = x^2 + 14y^2 \text{ or } 2x^2 + 7y^2 &\iff p \equiv 1,9,15,23,25,39 \bmod 56 \\ p = 3x^2 \pm 2xy + 5y^2 &\iff p \equiv 3,5,13,19,27,45 \bmod 56 \end{aligned} \tag{2.23}$$

when $p \neq 7$ is odd (see Exercise 2.15). Note that the top line proves part of Euler's conjecture (1.21) concerning $x^2 + 14y^2$.

In order to combine Theorem 2.16 and genus theory into a general theorem, we must show that the above examples reflect the general case. We first introduce some terminology. Given a negative integer $D \equiv 0,1 \bmod 4$, the *principal form* is defined to be

$$x^2 - \frac{D}{4}y^2, \qquad D \equiv 0 \bmod 4$$

$$x^2 + xy + \frac{1-D}{4}y^2, \qquad D \equiv 1 \bmod 4.$$

It is easy to check that the principal form has discriminant D and is reduced (see Exercise 2.16). Note that when $D = -4n$, we get our friend $x^2 + ny^2$. Using the principal form, we can characterize the congruence classes in $(\mathbb{Z}/D\mathbb{Z})^*$ represented by a form of discriminant D:

Lemma 2.24. *Given a negative integer $D \equiv 0,1 \bmod 4$, let $\ker(\chi) \subset (\mathbb{Z}/D\mathbb{Z})^*$ be as in Theorem 2.16, and let $f(x,y)$ be a form of discriminant D.*

(i) *The values in $(\mathbb{Z}/D\mathbb{Z})^*$ represented by the principal form of discriminant D form a subgroup $H \subset \ker(\chi)$.*

(ii) *The values in $(\mathbb{Z}/D\mathbb{Z})^*$ represented by $f(x,y)$ form a coset of H in $\ker(\chi)$.*

Proof. We first show that if a number m is prime to D and is represented by a form of discriminant D, then $[m] \in \ker(\chi)$. By Exercise 2.1, we can write $m = d^2m'$, where m' is properly represented by $f(x,y)$. Then $\chi([m]) = \chi([d^2m']) = \chi([d])^2\chi([m']) = \chi([m'])$. Thus we may assume that m is properly represented by $f(x,y)$, and then Lemma 2.5 implies that D is a quadratic residue modulo m, i.e., $D = b^2 - km$ for some b and k. When m is odd, the properties of the Jacobi symbol (see Lemma 1.14) imply that

$$\chi([m]) = \left(\frac{D}{m}\right) = \left(\frac{b^2 - km}{m}\right) = \left(\frac{b^2}{m}\right) = \left(\frac{b}{m}\right)^2 = 1$$

and our claim is proved. The case when m is even is covered in Exercise 2.17.

We now turn to statements (i) and (ii) of the lemma. Concerning (i), the above paragraph shows that $H \subset \ker(\chi)$. When $D = -4n$, the identity (1.6) shows that H is closed under multiplication, and hence H is a subgroup. When $D \equiv 1 \bmod 4$, the argument is slightly different: here, notice that

$$4\left(x^2+xy+\frac{1-D}{4}y^2\right) \equiv (2x+y)^2 \bmod D,$$

which makes it easy to show that H is in fact the subgroup of squares in $(\mathbb{Z}/D\mathbb{Z})^*$ (see Exercise 2.17).

To prove (ii), we need the following observation of Gauss [41, §228]:

Lemma 2.25. *Given a form $f(x,y)$ and an integer M, then $f(x,y)$ properly represents at least one number relatively prime to M.*

Proof. See Exercise 2.18. Q.E.D.

Now suppose that $D = -4n$. If we apply Lemma 2.25 with $M = 4n$ and then use Lemma 2.3, we may assume that $f(x,y) = ax^2 + bxy + cy^2$, where a is prime to $4n$. Since $f(x,y)$ has discriminant $-4n, b$ is even and can be written as $2b'$, and then (2.4) implies that

$$af(x,y) = (ax+b'y)^2+ny^2.$$

Since a is relatively prime to $4n$, it follows that the values of $f(x,y)$ in $(\mathbb{Z}/4n\mathbb{Z})^*$ lie in the coset $[a]^{-1}H$. Conversely, if $[c] \in [a]^{-1}H$, then $ac \equiv z^2+nw^2 \bmod 4n$ for some z and w. Using the above identity, it is easy to solve the congruence $f(x,y) \equiv c \bmod 4n$, and thus the coset $[a]^{-1}H$ consists exactly of the values represented in $(\mathbb{Z}/D\mathbb{Z})^*$ by $f(x,y)$. The case $D \equiv 1 \bmod 4$ is similar (see Exercise 2.17), and Lemma 2.24 is proved. Q.E.D.

Since distinct cosets of H are disjoint. Lemma 2.24 implies that different genera represent disjoint values in $(\mathbb{Z}/D\mathbb{Z})^*$. This allows us to describe genera by cosets H' of H in $\ker(\chi)$. We define the *genus of H'* to consist of all forms of discriminant D which represent the values of H' modulo D. Then Lemma 2.24 immediately implies the following refinement of Theorem 2.16:

Theorem 2.26. *Assume that $D \equiv 0, 1 \bmod 4$ is negative, and let $H \subset \ker(\chi)$ be as in Lemma 2.24. If H' is a coset of H in $\ker(\chi)$ and p is an odd prime not dividing D. then $[p] \in H'$ if and only if p is represented by a reduced form of discriminant D in the genus of H'.* Q.E.D.

This theorem is the main result of our elementary genus theory. It generalizes examples (2.22) and (2.23), and it shows that there are always congruence conditions which characterize when a prime is represented by some form in a given genus.

For us, the most interesting genus is the one containing the principal form, which following Gauss, we call the *principal genus.* When $D = -4n$, the principal form is $x^2 + ny^2$, and since $x^2 + ny^2$ is congruent modulo $4n$ to x^2 or $x^2 + n$, depending on

whether y is even or odd, we get the following explicit congruence conditions for this case:

Corollary 2.27. *Let n be a positive integer and p an odd prime not dividing n. Then p is represented by a form of discriminant $-4n$ in the principal genus if and only if for some integer β,*

$$p \equiv \beta^2 \text{ or } \beta^2 + n \bmod 4n. \qquad \text{Q.E.D.}$$

There is also a version of this for discriminants $D \equiv 1 \bmod 4$—see Exercise 2.20.

The nicest case of Corollary 2.27 is when the principal genus consists of a single class, for then we get congruence conditions that characterize $p = x^2 + ny^2$. This is what happened when $n = 5$ (see (2.22)), and this isn't the only case. For example, the table of reduced forms in Lagrange's memoir [69, pp. 766–767] shows that the same thing happens for $n = 6$, 10, 13, 15, 21, 22 and 30—for each of these n's, the principal genus consists of only one class (see Exercise 2.21). Corollary 2.27 then gives us the following theorems for primes p:

$$\begin{aligned}
p = x^2 + 6y^2 &\iff p \equiv 1,7 \bmod 24\\
p = x^2 + 10y^2 &\iff p \equiv 1,9,11,19 \bmod 40\\
p = x^2 + 13y^2 &\iff p \equiv 1,9,17,25,29,49 \bmod 52\\
p = x^2 + 15y^2 &\iff p \equiv 1,19,31,49 \bmod 60\\
p = x^2 + 21y^2 &\iff p \equiv 1,25,37 \bmod 84\\
p = x^2 + 22y^2 &\iff p \equiv 1,9,15,23,25,31,47,49,71,81 \bmod 88\\
p = x^2 + 30y^2 &\iff p \equiv 1,31,49,79 \bmod 120.
\end{aligned} \tag{2.28}$$

It should be clear that this is a powerful theory! A natural question to ask is how often does the principal genus consist of only one class, i.e., how many theorems like (2.28) do we get? We will explore this question in more detail in §3.

The genus theory just discussed has been very successful, but it hasn't solved all of the problems posed in §1. In particular, we have yet to prove Fermat's conjecture concerning $pq = x^2 + 5y^2$, and we've only done parts of Euler's conjectures (1.20) and (1.21) concerning $x^2 + 5y^2$ and $x^2 + 14y^2$. To complete the proofs, we again turn to Lagrange for help.

Let's begin with $x^2 + 5y^2$. We've already proved the part concerning when a prime p can equal $x^2 + 5y^2$ (see (2.22)), but it remains to show that for primes p and q, we have

$$\begin{aligned}
p,q \equiv 3,7 \bmod 20 &\implies pq = x^2 + 5y^2 \quad \text{(Fermat)}\\
p \equiv 3,7 \bmod 20 &\implies 2p = x^2 + 5y^2 \quad \text{(Euler)}.
\end{aligned} \tag{2.29}$$

Lagrange's argument [69, pp. 788–789] is as follows. He first notes that primes congruent to 3 or 7 modulo 20 can be written as $2x^2 + 2xy + 3y^2$ (this is (2.22)), so that both parts of (2.29) can be proved by showing that the product of two numbers

represented by $2x^2+2xy+3y^2$ is of the form x^2+5y^2. He then states the identity

$$(2.30)\qquad \begin{aligned}(2x^2+2xy+3y^2)(2z^2+2zw+3w^2)\\ =(2xz+xw+yz+3yw)^2+5(xw-yz)^2\end{aligned}$$

(see Exercise 2.22), and everything is proved!

Turning to Euler's conjecture (1.21) for x^2+14y^2, we proved part of it in (2.23), but we still need to show that

$$p\equiv 3,5,13,19,27,45 \bmod 56 \iff 3p=x^2+14y^2.$$

Using (2.23), it suffices to show that 3 times a number represented by $3x^2\pm 2xy+5y^2$, or more generally the product of any two such numbers, is of the form x^2+14y^2. So what we need is another identity of the form (2.30), and in fact there is a version of (2.30) that holds for any form of discriminant $-4n$:

$$(2.31)\qquad \begin{aligned}(ax^2+2bxy+cy^2)(az^2+2bzw+cw^2)\\ =(axz+bxw+byz+cyw)^2+n(xw-yz)^2\end{aligned}$$

(see Exercise 2.21). Applying this to $3x^2+2xy+5y^2$ and $n=14$, we are done.

We can also explain one other aspect of Euler's conjectures (1.20) and (1.21), for recall that we wondered why (1.20) used $2p$ while (1.21) used $3p$. The answer again involves the identities (2.30) and (2.31): they show that 2 (resp. 3) can be replaced by any value represented by $2x^2+2xy+3y^2$ (resp. $3x^2\pm 2xy+5y^2$). But Legendre's observation from the proof of Theorem 2.8 shows that 2 (resp. 3) is the best choice because it's the smallest nonzero value represented by the form in question. We will see below and in §3 that identities like (2.30) and (2.31) are special cases of the composition of quadratic forms.

We now have complete proofs of Euler's conjectures (1.20) and (1.21) for x^2+5y^2 and x^2+14y^2. Notice that we've used a lot of mathematics: quadratic reciprocity, reduced quadratic forms, genus theory and the composition of quadratic forms. This amply justifies the high estimate of Euler's insight that was made in §1, and Lagrange is equally impressive for providing the proper tools to understand what lay behind Euler's conjectures.

D. Lagrange and Legendre

We've already described parts of Lagrange's memoir "Recherches d'Arithmétique," but there are some further comments we'd like to add. First, although we credit Lagrange with the discovery of genus theory, it appears only implicitly in his work. The groupings that appear in his tables of reduced forms are striking, but Lagrange's comments on genus theory are a different matter. On the page before the tables begin, Lagrange explains his grouping of forms as follows: "when two different [forms] give the same values of b [in $(\mathbb{Z}/4n\mathbb{Z})^*$], one combines these [forms] into the same case" [69, p. 765]. This is the sum total of what Lagrange says about genus theory!

After completing the basic theory of quadratic forms (both definite and indefinite), Lagrange gives some applications to number theory. To motivate his results, he turns to Fermat and Euler, and he quotes from two of our main sources of inspiration: Fermat's 1658 letter to Digby and Euler's 1744 paper on prime divisors of $paa \pm qyy$. Lagrange explicitly states Fermat's results (1.1) on primes of the form x^2+ny^2, $n = 1, 2$ or 3, and he notes Fermat's speculation that $pq = x^2+5y^2$ whenever p and q are primes congruent to 3 or 7 modulo 20. Lagrange also mentions several of Euler's conjectures, including (1.20), and he adds "one finds a very large number of similar theorems in Volume XIV of the old *Commentaires de Pétersbourg* [where Euler's 1744 paper appeared], but none of them have been demonstrated until now" [69, pp. 775–776].

The last section of Lagrange's memoir is titled "Prime numbers of the form $4nm + b$ which are at the same time of the form $x^2 \pm ny^2$" [69, p. 775]. It's clear that Lagrange wanted to prove Theorem 2.26, so that he could read off corollaries like (2.17), (2.22), (2.23) and (2.28). The problem is that these proofs depend on quadratic reciprocity, which Lagrange didn't know in general—he could only prove some special cases. For example, he was able to determine $(\pm 2/p)$, $(\pm 3/p)$ and $(\pm 5/p)$, but he had only partial results for $(\pm 7/p)$. Thus, he could prove all of (2.22) but only parts of the others (see [69, pp. 784–793] for the full list of his results). To get the flavor of Lagrange's arguments, the reader should see Exercise 2.23 or Scharlau and Opolka [86, pp. 41–43]. At the end of the memoir, Lagrange summarizes what he could prove about quadratic reciprocity, stating his results in terms of Euler's criterion

$$a^{(p-1)/2} \equiv \left(\frac{a}{p}\right) \bmod p.$$

For example, for $(2/p)$, Lagrange states [69, p. 794]:

> Thus, if p is a prime number of one of the forms $8n \pm 1$, $2^{(p-1)/2} - 1$ will be divisible by p, and if p is of the form $8n \pm 3, 2^{(p-1)/2} + 1$ will thus be divisible by p.

We next turn to Legendre. In his 1785 memoir "Recherches d'Analyse Indéterminée" [75], the two major results are first, a necessary and sufficient criterion for the equation

$$ax^2 + by^2 + cz^2 = 0, \qquad a, b, c \in \mathbb{Z}$$

to have a nontrivial integral solution, and second, a proof of quadratic reciprocity. Legendre was influenced by Lagrange, but he replaces Lagrange's "$2^{(p-1)/2} - 1$ will be divisible by p" by the simpler phrase "$2^{(p-1)/2} = 1$," where, as he warns the reader, "one has thrown out the multiples of p in the first member" [75, p. 516]. He then goes on to state quadratic reciprocity in the following form [75, p. 517]:

> c and d being two [odd] prime numbers, the expressions $c^{(d-1)/2}$, $d^{(c-1)/2}$ do not have different signs except when c & d are both of the form $4n - 1$; in all other cases, these expressions will always have the same sign.

Except for the notation, this is a thoroughly modern statement of quadratic reciprocity. Legendre's proof is a different matter, for it is quite incomplete. We won't examine the proof in detail—this is done in Weil [106, pp. 328–330 and 344–345].

Suffice it to say that some of the cases are proved rigorously (see Exercise 2.24), some depend on Dirichlet's theorem on primes in arithmetic progressions, and some are a tangle of circular reasoning.

In 1798 Legendre published a more ambitious work, the *Essai sur la Théorie des Nombres.* (The third edition [74], published 1830, was titled *Théorie des Nombres,* and all of our references will be to this edition.) Legendre must have been dissatisfied with the notation of the "Recherches", for in the *Essai* he introduces the Legendre symbol (a/p). Then, in a section titled "Theorem containing a law of reciprocity which exists between two arbitrary prime numbers," Legendre states that if n and m are distinct odd primes, then

$$\left(\frac{n}{m}\right) = (-1)^{(n-1)/2\cdot(m-1)/2}\left(\frac{m}{n}\right)$$

(see [74, Vol. I, p. 230]). This is where our notation and terminology for quadratic reciprocity come from. Unfortunately, the *Essai* repeats Legendre's incomplete proof from 1785, although by the 1830 edition there had been enough criticism of this proof that Legendre added Gauss' third proof of reciprocity as well as one communicated to him by Jacobi (still maintaining that his original proof was valid).

The *Essai* also contains a treatment of quadratic forms. Like Lagrange, one of Legendre's goals was to prove theorems in number theory using quadratic forms. The difference is that Legendre knows quadratic reciprocity (or at least he thinks he does), and this allows him to state a version of our main result, Theorem 2.26. Legendre calls it his "Théorème General" [74, Vol. I, p. 299], and it goes as follows: if $[a]$ is a congruence class lying in $\ker(\chi)$, then

> every prime number comprised of the form $4nx + a\ldots$ will consequently be given by one of the quadratic forms $py^2 + 2qyz \pm rz^2$ which correspond to the linear form $4nx + a$.

The terminology here is interesting. Euler and Lagrange would speak of numbers "of the form" $4nx + a$ or "of the form" $ax^2 + bxy + cy^2$. As the above quote indicates, Legendre distinguished these two by calling them linear forms and quadratic forms respectively. This is where we get the term "quadratic form."

While Legendre's "Théorème" makes no explicit reference to genus theory, the context shows that it's there implicitly. Namely, Legendre's book has tables similar to Lagrange's, with the forms grouped according to the values they represent in $(\mathbb{Z}/D\mathbb{Z})^*$. Since the explanation of the tables immediately precedes the statement of the "Théorème" [74, Vol. I, pp. 286–298], it's clear that Legendre's correspondence between linear forms and quadratic forms is exactly that given by Theorem 2.26.

To Legendre, this theorem "is, without contradiction, one of the most general and most important in the theory of numbers" [74, Vol. I, p. 302]. Its main consequence is that every entry in his tables becomes a theorem, and Legendre gives several pages of explicit examples [74, Vol. I, pp. 305–307]. This is a big advance over what Lagrange could do, and Legendre notes that quadratic reciprocity was the key to his success [74, Vol. I, p. 307]:

> Lagrange is the first who opened the way for the study of these sorts of theorems. ... But the methods which served the great geometer are not applicable ...

except in very few cases; and the difficulty in this regard could not be completely resolved without the aid of the law of reciprocity.

Besides completing Lagrange's program, Legendre also tried to understand some of the other ideas implicit in Lagrange's memoir. We will discuss one of Legendre's attempts that is particularly relevant to our purposes: his theory of composition. Legendre's basic idea was to generalize the identity (2.30)

$$\begin{aligned}(2x^2+2xy+3y^2)(2z^2+2zw+3w^2)\\ =(2xz+xw+yz+3yw)^2+5(xw-yz)^2\end{aligned}$$

used by Lagrange in proving the conjectures of Fermat and Euler concerning x^2+5y^2. We gave one generalization in (2.31), but Legendre saw that something more general was going on. More precisely, let $f(x,y)$ and $g(x,y)$ be forms of discriminant D. Then a form $F(x,y)$ of the same discriminant is their *composition* provided that

$$f(x,y)g(z,w) = F(B_1(x,y;z,w),B_2(x,y;z,w))$$

where

$$B_i(x,y;z,w) = a_ixz+b_ixw+c_iyz+d_iyw, \qquad i=1,2$$

are bilinear forms in x,y and z,w. Thus Lagrange's identity shows that x^2+5y^2 is the composition of $2x^2+2xy+3y^2$ with itself. And this is not the only example we've seen—the reader can check that (1.3), (1.6) and (2.31) are also examples of the composition of forms.

A useful consequence of composition is that whenever $F(x,y)$ is composed of $f(x,y)$ and $g(x,y)$, then the product of numbers represented by $f(x,y)$ and $g(x,y)$ will be represented by $F(x,y)$. This was the idea that enabled us to complete the conjectures of Fermat and Euler for x^2+5y^2 and x^2+14y^2.

The basic question is whether any two forms of the same discriminant can be composed, and Legendre showed that the answer is yes [74, Vol. II, pp. 27–30]. For simplicitly, let's discuss the case where the forms $f(x,y)=ax^2+2bxy+cy^2$ and $g(x,y)=a'x^2+2b'xy+c'y^2$ have discriminant $-4n$, and a and a' are relatively prime (we can always arrange the last condition by changing the forms by a proper equivalence). Then the Chinese Remainder Theorem shows that there is a number B such that

$$\begin{aligned}B &= \pm b \bmod a\\ B &= \pm b' \bmod a'.\end{aligned} \tag{2.32}$$

It follows that $B^2+n \equiv b^2+(ac-b^2) \equiv 0 \bmod a$, so that $a \mid B^2+n$. The same holds for a', and thus $aa' \mid B^2+n$. Then Legendre shows that the form

$$F(x,y) = aa'x^2+2Bxy+\frac{B^2+n}{aa'}y^2$$

is the composition of $f(x,y)$ and $g(x,y)$. A modern account of Legendre's argument may be found in Weil [106, pp. 332–335]), and we will consider this problem (from a slightly different point of view) in §3 when we discuss composition in more detail.

Because of the $\pm$ signs in (2.32), two forms in general may be composed in four different ways. For example, the forms $14x^2+10xy+21y^2$ and $9x^2+2xy+30y^2$ compose to the four forms

$$126x^2 \pm 38xy + 5y^2, \qquad 126x^2 \pm 74xy + 13y^2,$$

and it is easy to show that these forms all lie in different classes (see Exercise 2.26). Since Legendre used equivalence rather than proper equivalence, he sees two rather than four forms here—for him, this operation "leads in general to two solutions" [74, Vol. II, p. 28].

One of Legendre's important ideas is that since every form is equivalent to a reduced one, it suffices to work out the compositions of reduced forms. The resulting table would then give the compositions of all possible forms of that discriminant. Let's look at the case $n = 41$, which Legendre does in detail in [74, Vol. II, pp. 39–40]. He labels the reduced forms as follows:

$$\begin{aligned} A &= x^2+41y^2 \\ B &= 2x^2+2xy+21y^2 \\ C &= 5x^2+4xy+9y^2 \\ D &= 3x^2+2xy+14y^2 \\ E &= 6x^2+2xy+7y^2. \end{aligned} \tag{2.33}$$

(Legendre writes the forms slightly differently, but it's more convenient to work with reduced forms.) He then gives the following table of compositions:

(2.34)	$AA = A$	$BB = A$	$CC = A$ or B	$DD = A$ or C	$EE = A$ or C
	$AB = B$	$BC = C$	$CD = D$ or E	$DE = B$ or C	
	$AC = C$	$BD = E$	$CE = D$ or E		
	$AD = D$	$BE = D$			
	$AE = E$				

This almost looks like the multiplication table for a group, but the binary operation isn't single-valued. To the modern reader, it's clear that Legendre must be doing something slightly wrong.

One problem is that (2.33) lists 5 forms, while the class number is 8. (C, D and E each give two reduced forms, while A and B each give only one.) This is closely related to the ambiguity in Legendre's operation: as long as we work with equivalence rather than proper equivalence, we can't fix the sign of the middle coefficient $2b$ of a reduced form, so that the $\pm$ signs in (2.32) are forced upon us.

This suggests that composition might give a group operation on the *classes* of forms of discriminant D. However, there remain serious problems to be solved. Composition, as defined above, is still a multiple-valued operation. Thus one has to show that the signs in (2.32) can be chosen *uniformly* so that as we vary $f(x,y)$ and $g(x,y)$ within their proper equivalance classes, the resulting compositions are all properly equivalent. Then one has to worry about associativity, inverses, etc. There's a lot of work to be done!

This concludes our discussion of Lagrange and Legendre. While the last few pages have raised more questions than answers, the reader should still be convinced of the richness of the theory of quadratic forms. The surprising fact is that we have barely reached the really interesting part of the theory, for we have yet to consider the work of Gauss.

E. Exercises

2.1. If a form $f(x,y)$ represents an integer m, show that m can be written $m = d^2m'$, where $f(x,y)$ properly represents m'.

2.2. In this exercise we study equivalence and proper equivalence.

(a) Show that equivalence and proper equivalence are equivalence relations.

(b) Show that improper equivalence is not an equivalence relation.

(c) Show that equivalent forms represent the same numbers, and show that the same holds for proper representations.

(d) Show that any form equivalent to a primitive form is itself primitive. Hint: use (c).

2.3. Let $f(x,y)$ and $g(x,y)$ be forms of discriminants D and D' respectively, and assume that there are integers p,q,r and s such that

$$f(x,y) = g(px+qy, rx+sy).$$

Prove that $D = (ps-qr)^2D'$.

2.4. Let $f(x,y)$ be a form of discriminant $D \neq 0$.

(a) If $D > 0$, then use (2.4) to prove that $f(x,y)$ represents both positive and negative numbers.

(b) If $D < 0$, then show that $f(x,y)$ represents only positive or only negative numbers, depending on the sign of the coefficient of x^2.

2.5. Formulate and prove a version of Corollary 2.6 which holds for arbitrary discriminants.

2.6. Find a reduced form that is properly equivalent to $126x^2+74xy+13y^2$. Hint: make the middle coefficient small—see the proof of Theorem 2.8.

2.7. Prove (2.9) for forms that satisfy $|b| \leq a \leq c$.

2.8. This exercise is concerned with the uniqueness part of Theorem 2.8.

(a) Prove (2.11).

(b) Prove a version of (2.11) that holds in the exceptional cases $|b| = a$ or $a = c$, and use this to complete the uniqueness part of the proof of Theorem 2.8.

2.9. Use a computer algebra system (such as *Maple* or *Mathematica*) to write a procedure that computes the class number and all reduced forms of a given discriminant $D < 0$. For example, one finds that $h(-32767) = 52$. If you don't use a computer, then you should check the following examples by hand.

(a) Verify the entries in table (2.14).

(b) Compute all reduced forms of discriminants -3, -15, -24, -31 and -52.

2.10. This exercise is concerned with indefinite forms of discriminant $D > 0$, D not a perfect square. The last condition implies that the outer coefficients of a form with discriminant D are nonzero.

(a) Adapt the proof of Theorem 2.8 to show that any form of discriminant D is properly equivalent to $ax^2 + bxy + cy^2$, where

$$|b| \leq |a| \leq |c|.$$

(b) If $ax^2 + bxy + cy^2$ satifies the above inequalities, prove that

$$|a| \leq \frac{\sqrt{D}}{2}.$$

(c) Conclude that there are only finitely many proper equivalence classes of forms of discriminant D. This proves that the class number $h(D)$ is finite.

2.11. Use Theorem 2.16, quadratic reciprocity and table (2.14) to prove Fermat's three theorems (1.1) and the new result (2.17) for $x^2 + 7y^2$.

2.12. This exercise is concerned with the proof of Theorem 2.18.

(a) If $m > 1$ is an integer which is not a prime power, prove that m can be written $m = ac$ where $1 < a < c$ and $\gcd(a, c) = 1$.

(b) Show that $h(-32) = 2$ and $h(-124) = 3$.

2.13. Use Theorem 2.16, quadratic reciprocity and table (2.14) to prove (2.19), and work out similar results for discriminants -3, -15, -24, -31 and -52.

2.14. Prove (2.20) and (2.21). Hint: use Lemma 2.24.

2.15. Prove (2.23).

2.16. Let D be a number congruent to 1 modulo 4. Show that the form $x^2 + xy + ((1 - D)/4)y^2$ has discriminant D, and show that it is reduced when $D < 0$.

2.17. In this exercise, we will complete the proof of Lemma 2.24 for discriminants $D \equiv 1 \bmod 4$. Let $\chi : (\mathbb{Z}/D\mathbb{Z})^* \to \{\pm 1\}$ be as in Lemma 1.14.

(a) If an even number is properly represented by a form of discriminant D, then show that $D \equiv 1 \bmod 8$. Hint: use Lemma 2.3.

(b) If m is relatively prime to D and is represented by a form of discriminant D, then show that $[m] \in \ker(\chi)$. Hint: use Lemma 2.5 and, when m is even, (a) and Exercise 1.12(c).

(c) Let $H \subset (\mathbb{Z}/D\mathbb{Z})^*$ be the subgroup of squares. Show that H consists of the values represented by $x^2 + xy + ((1-D)/4)y^2$. Hint: use

$$4\left(x^2 + xy + \frac{1-D}{4}y^2\right) \equiv (2x+y)^2 \bmod D.$$

(d) Let $f(x,y)$ be a form of discriminant D. Show that the values in $(\mathbb{Z}/D\mathbb{Z})^*$ represented by $f(x,y)$ form a coset of H in $\ker(\chi)$. Hint: use (2.4).

2.18. Let $f(x,y) = ax^2 + bxy + cy^2$, where as usual we assume $\gcd(a,b,c) = 1$.

(a) Given a prime p, prove that at least one of $f(1,0)$, $f(0,1)$ and $f(1,1)$ is relatively prime to p.

(b) Prove Lemma 2.25. Hint: use (a) and the Chinese Remainder Theorem.

2.19. Work out the genus theory of Theorem 2.26 for discriminants -15, -24, -31 and -52. Your answers should be similar to (2.22) and (2.23).

2.20. Formulate and prove a version of Corollary 2.27 for negative discriminants $D \equiv 1 \bmod 4$. Hint: by Exercise 2.17(c), H is the subgroup of squares.

2.21. Prove (2.28). Hint: for each n, find the reduced forms and use Lemma 2.24.

2.22. Prove (2.30) and its generalization (2.31).

2.23. The goal of this exercise is to prove that $(-2/p) = 1$ when $p \equiv 1,3 \bmod 8$. The argument below is due to Lagrange, and is similar to the one used by Euler in his proof of the Reciprocity Step for $x^2 + 2y^2$ [33, Vol. II, pp. 240–281].

(a) When $p \equiv 1 \bmod 8$, write $p = 8k+1$, and then use the identity

$$x^{8k} - 1 = ((x^{2k} - 1)^2 + 2x^{2k})(x^{4k} - 1)$$

to show that $(-2/p) = 1$.

(b) When $p \equiv 3 \bmod 8$, assume that $(-2/p) = -1$. Show that $(2/p) = 1$, and thus by Corollary 2.6, p is represented by a form of discriminant 8.

(c) Use Exercise 2.10(a) to show that any form of discriminant 8 is properly equivalent to $\pm(x^2 - 2y^2)$.

(d) Show that an odd prime $p = \pm(x^2 - 2y^2)$ is congruent to ± 1 modulo 8.

From (a)–(d), it follows easily that $(-2/p) = 1$ when $p \equiv 1, 3 \bmod 8$.

2.24. One of the main theorems in Legendre's 1785 memoir [74, pp. 509–513] states that the equation

$$ax^2 + by^2 + cz^2 = 0,$$

where abc is squarefree, has a nontrivial integral solution if and only if

(i) a, b and c are not all of the same sign, and

(ii) $-bc$, $-ac$ and $-ab$ are quadratic residues modulo $|a|$, $|b|$ and $|c|$ respectively.

As we've already noted, Legendre tried to use this result to prove quadratic reciprocity. In this problem, we will treat one of the cases where he succeeded. Let p and q be primes which satisfy $p \equiv 1 \bmod 4$ and $q \equiv 3 \bmod 4$, and assume that $(p/q) = -1$ and $(q/p) = 1$. We will derive a contradiction as follows:

(a) Use Legendre's theorem to show that $x^2 + py^2 - qz^2 = 0$ has a nontrivial integral solution.

(b) Working modulo 4, show that $x^2 + py^2 - qz^2 = 0$ has no nontrivial integral solutions.

In [106, pp. 339–345], Weil explains why this argument works.

2.25. The opposite of the form $ax^2 + bxy + cy^2$ is the form $ax^2 - bxy + cy^2$. Prove that two forms are properly equivalent if and only if their opposites are.

2.26. Verify that $14x^2 + 10xy + 21y^2$ and $9x^2 + 2xy + 30y^2$ compose to the four forms $126x^2 \pm 74xy + 13y^2$ and $126x^2 \pm 38xy + 5y^2$, and show that they all lie in different classes. Hint: use Exercises 2.6 and 2.25.

2.27. Let p be a prime number which is represented by forms $f(x,y)$ and $g(x,y)$ of the same discriminant.

(a) Show that $f(x,y)$ and $g(x,y)$ are equivalent. Hint: use Lemma 2.3, and examine the middle coefficient modulo p.

(b) If $f(x,y) = x^2 + ny^2$, and $g(x,y)$ is reduced, then show that $f(x,y)$ and $g(x,y)$ are equal.

§3. GAUSS, COMPOSITION AND GENERA

While genus theory and composition were implicit in Lagrange's work, these concepts are still primarily linked to Gauss, and for good reason: he may not have been the first to use them, but he was the first to understand their astonishing depth and interconnection. In this section we will prove Gauss' major results on composition

and genus theory for the special case of positive definite forms. We will then apply this theory to our question concerning primes of the form x^2+ny^2, and we will also discuss Euler's convenient numbers. These turn out to be those n's for which each genus consists of a single class, and it is still not known exactly how many there are. The section will end with a discussion of Gauss' *Disquisitiones Arithmeticae.*

A. Composition and the Class Group

The basic definition of composition was given in §2: if $f(x,y)$ and $g(x,y)$ are primitive positive definite forms of discriminant D, then a form $F(x,y)$ of the same type is their *composition* provided that

$$f(x,y)g(z,w) = F(B_1(x,y;z,w), B_2(x,y;z,w)),$$

where

$$B_i(x,y;z,w) = a_ixz + b_ixw + c_iyz + d_iyw, \qquad i = 1,2$$

are integral bilinear forms. Two forms can be composed in many different ways, and the resulting forms need not be properly equivalent. In §2 we gave an example of two forms whose compositions lay in four distinct classes. So if we want a well-defined operation on classes of forms, we must somehow *restrict* the notion of composition. Gauss does this as follows: given the above composition data, he proves that

$$a_1b_2 - a_2b_1 = \pm f(1,0), \qquad a_1c_2 - a_2c_1 = \pm g(1,0) \tag{3.1}$$

(see [41, §235] or Exercise 3.1), and then he defines the composition to be a *direct composition* provided that both of the signs in (3.1) are $+$.

The main result of Gauss' theory of composition is that for a fixed discriminant, direct composition makes the set of classes of forms into a finite Abelian group [41, §§236–240, 245 and 249]. Unfortunately, direct composition is an awkward concept to work with, and Gauss' proof of the group structure is long and complicated. So rather than follow Gauss, we will take a different approach to the study of composition. The basic idea is due to Dirichlet [28, Supplement X], though his treatment was clearly influenced by Legendre. Before giving Dirichlet's definition, we will need the following lemma:

Lemma 3.2. *Assume that* $f(x,y) = ax^2+bxy+cy^2$ *and* $g(x,y) = a'x^2+b'xy+c'y^2$ *have discriminant* D *and satisfy* $\gcd(a,a',(b+b')/2) = 1$ *(since* b *and* b' *have the same parity,* $(b+b')/2$ *is an integer). Then there is a unique integer* B *modulo* $2aa'$ *such that*

$$\begin{aligned} B &\equiv b \bmod 2a \\ B &\equiv b' \bmod 2a' \\ B^2 &\equiv D \bmod 4aa'. \end{aligned}$$

Proof. The first step is to put these congruences into a standard form. If a number B satisfies the first two, then

$$B^2-(b+b')B+bb' \equiv (B-b)(B-b') \equiv 0 \bmod 4aa',$$

so that the third congruence can be written as

$$(b+b')B \equiv bb'+D \bmod 4aa'.$$

Dividing by 2, this becomes

$$(b+b')/2 \cdot B \equiv (bb'+D)/2 \bmod 2aa'. \tag{3.3}$$

If we multiply the first two congruences of the lemma by a' and a respectively and combine them with (3.3), we see that the three congruences in the statement of the lemma are equivalent to

$$\begin{aligned} a' \cdot B &\equiv a'b \bmod 2aa' \\ a \cdot B &\equiv ab' \bmod 2aa' \\ (b+b')/2 \cdot B &\equiv (bb'+D)/2 \bmod 2aa'. \end{aligned} \tag{3.4}$$

The following lemma tells us about the solvability of these congruences:

Lemma 3.5. *Let $p_1,q_1,\ldots,p_r,q_r,m$ be numbers with $\gcd(p_1,\ldots,p_r,m)=1$. Then the congruences*

$$p_iB \equiv q_i \bmod m, \qquad i=1,\ldots,r$$

have a unique solution modulo m if and only if for all $i,j=1,\ldots,r$ we have

$$p_iq_j \equiv p_jq_i \bmod m. \tag{3.6}$$

Proof. See Exercise 3.3. Q.E.D.

Since we are assuming $\gcd(a,a',(b+b')/2)=1$, the congruences (3.4) satisfy the gcd condition of the above lemma, and the compatibility conditions (3.6) are easy to verify (see Exercise 3.4). The existence and uniqueness of the desired B follow immediately. Q.E.D.

We now give Dirichlet's definition of composition. Let $f(x,y)=ax^2+bxy+cy^2$ and $g(x,y)=a'x^2+b'xy+c'y^2$ be primitive positive definite forms of discriminant $D<0$ which satisfy $\gcd(a,a',(b+b')/2)=1$. Then the *Dirichlet composition* of $f(x,y)$ and $g(x,y)$ is the form

$$F(x,y)=aa'x^2+Bxy+\frac{B^2-D}{4aa'}y^2, \tag{3.7}$$

where B is the integer determined by Lemma 3.2. The basic properties of $F(x,y)$ are:

Proposition 3.8. *Let $f(x,y)$ and $g(x,y)$ be as above. Then the Dirichlet composition $F(x,y)$ defined in* (3.7) *is a primitive positive definite form of discriminant D, and $F(x,y)$ is the direct composition of $f(x,y)$ and $g(x,y)$ in the sense of* (3.1).

Proof. An easy calculation shows that $F(x,y)$ has discriminant D, and the form is consequently positive definite.

The next step is to prove that $F(x,y)$ is the composition of $f(x,y)$ and $g(x,y)$. We will sketch here the argument and leave the details to the reader. To begin, let $C = (B^2 - D)/4aa'$, so that $F(x,y) = aa'x^2 + Bxy + Cy^2$. Then, using the first two congruences of Lemma 3.2, it is easy to prove that $f(x,y)$ and $g(x,y)$ are properly equivalent to the forms $ax^2 + Bxy + a'Cy^2$ and $a'x^2 + Bxy + aCy^2$ respectively. However, for these last two forms one has the composition identity

$$(ax^2 + Bxy + a'Cy^2)(a'z^2 + Bzw + aCw^2) = aa'X^2 + BXY + CY^2,$$

where $X = xz - Cyw$ and $Y = axw + a'yz + Byw$. It follows easily that $F(x,y)$ is the composition of $f(x,y)$ and $g(x,y)$. With a little more effort, it can be checked that this is a direct composition in Gauss' sense (3.1). The details of these arguments are covered in Exercise 3.5.

It remains to show that $F(x,y)$ is primitive, i.e., that its coefficients are relatively prime. Suppose that some prime p divided all of the coefficients. This would imply that p divided all numbers represented by $F(x,y)$. Since $F(x,y)$ is the composition of $f(x,y)$ and $g(x,y)$, this implies that p divides all numbers of the form $f(x,y)g(z,w)$. But $f(x,y)$ and $g(x,y)$ are primitive, so that by Lemma 2.25, they represent numbers relatively prime to p. Hence $f(x,y)g(z,w)$ also represents a number relatively prime to p. This contradiction completes the proof of the proposition. Q.E.D.

While Dirichlet composition is not as general as direct composition (not all direct compositions satisfy $\gcd(a,a',(b+b')/2) = 1$), it is easier to use in practice since there is an explicit formula (3.7) for the composition. Notice also that the congruence conditions in Lemma 3.2 are similar to the ones (2.32) used by Legendre. This is no accident, for when $D = -4n$ and $\gcd(a,a') = 1$, Dirichlet's formula reduces exactly to the one given by Legendre (see Exercise 3.6).

We can now state our main result on composition:

Theorem 3.9. *Let $D \equiv 0,1 \bmod 4$ be negative, and let $C(D)$ be the set of classes of primitive positive definite forms of discriminant D. Then Dirichlet composition induces a well-defined binary operation on $C(D)$ which makes $C(D)$ into a finite Abelian group whose order is the class number $h(D)$.*

Furthermore, the identity element of $C(D)$ is the class containing the principal form

$$x^2 - \frac{D}{4}y^2 \qquad \text{if } D \equiv 0 \bmod 4$$

$$x^2 + xy + \frac{1-D}{4}y^2 \qquad \text{if } D \equiv 1 \bmod 4,$$

and the inverse of the class containing the form $ax^2 + bxy + cy^2$ is the class containing $ax^2 - bxy + cy^2$.

Remarks. Some terminology is in order here.

(i) The group $C(D)$ is called the *class group,* though we will sometimes refer to $C(D)$ as the *form class group* to distinguish it from the ideal class group to be defined later.

(ii) The principal form of discriminant D was introduced in §2. The class it lies in is called the *principal class.* When $D = -4n$, the principal form is $x^2 + ny^2$.

(iii) The form $ax^2 - bxy + cy^2$ is called the *opposite* of $ax^2 + bxy + cy^2$, so that the opposite form gives the inverse under Dirichlet composition.

Proof. Let $f(x,y) = ax^2 + bxy + cy^2$ and $g(x,y)$ be forms of the given type. Using Lemmas 2.3 and 2.25, we can replace $g(x,y)$ by a properly equivalent form $a'x^2 + b'xy + c'y^2$ where $\gcd(a,a') = 1$. Then the Dirichlet composition of these forms is defined, which proves that Dirichlet composition is defined for any pair of classes in $C(D)$. To get a group structure out of this, we must then prove that:

(i) This operation is well-defined on the level of classes, and

(ii) The induced binary operation makes $C(D)$ into an Abelian group.

The proofs of (i) and (ii) can be done directly using the definition of Dirichlet composition (see Dirichlet [28, Supplement X] or Flath [36, §V.2]), but the argument is much easier using ideal class groups (to be studied in §7). We will therefore postpone this part of the proof until then. For now, we will assume that (i) and (ii) are true.

Let's next show that the principal class is the identity element of $C(D)$. To compose the principal form with $f(x,y) = ax^2 + bxy + cy^2$, first note that the gcd condition is clearly met, and thus the Dirichlet composition is defined. Then observe that $B = b$ satisfies the conditions of Lemma 3.2, so that by formula (3.7), the Dirichlet composition $F(x,y)$ reduces to the given form $f(x,y)$. This proves that the principal class is the identity.

Finally, given $f(x,y) = ax^2 + bxy + cy^2$, its opposite is $f'(x,y) = ax^2 - bxy + cy^2$. Since $\gcd(a,a,(b+(-b))/2) = a$ may be > 1, we can't apply Dirichlet composition directly. But if we use the proper equivalence $(x,y) \mapsto (-y,x)$, then we can replace $f'(x,y)$ by $g(x,y) = cx^2 + bxy + ay^2$. Since $\gcd(a,c,(b+b)/2) = \gcd(a,c,b) = 1$, we can apply Dirichlet's formulas to $f(x,y)$ and $g(x,y)$. One checks easily that $B = b$ satisfies the conditions of Lemma 3.2, so that the Dirichlet composition is $acx^2 + bxy + y^2$. We leave it to the reader to show that this form is properly equivalent to the principal form (see Exercise 3.7). This completes the proof of the theorem. Q.E.D.

We can now complete the discussion (begun in §2) of Legendre's theory of composition. To prevent confusion, we will distinguish between a *class* (all forms properly equivalent to a given form) and a *Lagrangian class* (all forms equivalent to a given one). In Theorem 3.9, we studied the composition of classes, while Legendre was concerned with the composition of Lagrangian classes. It is an easy exercise to show that the Lagrangian class of a form is the union of its class and the class of its

opposite (see Exercise 3.8). Theorem 3.9 implies that a Lagrangian class is the union of a class and its inverse in the class group $C(D)$. Thus Legendre's "operation" is the multiple-valued operation that multiplication induces on the set $C(D)/\sim$, where $\sim$ is the equivalence relation that identifies $x \in C(D)$ with x^{-1} (see Exercise 3.9). In Legendre's example (2.33), which dealt with forms of discriminant -164, we will see shortly that $C(-164) \simeq \mathbb{Z}/8\mathbb{Z}$, and it is then an easy exercise to show that $C(-164)/\sim$ is isomorphic to the structure given in (2.34) (see Exercise 3.9).

Elements of order ≤ 2 in the class group $C(D)$ play a special role in composition and genus theory. The reduced forms that lie in such classes are easy to find:

Lemma 3.10. *A reduced form $f(x,y) = ax^2 + bxy + cy^2$ of discriminant D has order ≤ 2 in the class group $C(D)$ if and only if $b = 0$, $a = b$ or $a = c$.*

Proof. Let $f'(x,y)$ be the opposite of $f(x,y)$. By Theorem 3.9, the class of $f(x,y)$ has order ≤ 2 if and only if the forms $f(x,y)$ and $f'(x,y)$ are properly equivalent. There are two cases to consider:

$|b| < a < c$: Here, $f'(x,y)$ is also reduced, so that by Theorem 2.8, the two forms are properly equivalent $\iff b = 0$.

$a = b$ or $a = c$: In these cases, the proof of Theorem 2.8 shows that the two forms are always properly equivalent.

The lemma now follows immediately. Q.E.D.

For an example of how this works, consider Legendre's example from §2 of forms of discriminant -164. The reduced forms are listed in (2.33), and Lemma 3.10 shows that only $2x^2 + 2xy + 21y^2$ has order 2. Since the class number is 8, the structure theorem for finite Abelian groups shows that the class group $C(-164)$ must be $\mathbb{Z}/8\mathbb{Z}$.

A surprising fact is that one doesn't need to list the reduced forms in order to determine the number of elements of order 2 in the class group:

Proposition 3.11. *Let $D \equiv 0, 1 \bmod 4$ be negative, and let r be the number of odd primes dividing D. Define the number μ as follows: if $D \equiv 1 \bmod 4$, then $\mu = r$, and if $D \equiv 0 \bmod 4$, then $D = -4n$, where $n > 0$, and μ is determined by the following table:*

n	μ
$n \equiv 3 \bmod 4$	r
$n \equiv 1, 2 \bmod 4$	$r+1$
$n \equiv 4 \bmod 8$	$r+1$
$n \equiv 0 \bmod 8$	$r+2$

Then the class group $C(D)$ has exactly $2^{\mu-1}$ elements of order ≤ 2.

Proof. For simplicity, we will treat only the case $D = -4n$ where $n \equiv 1 \bmod 4$. Recall that a form of discriminant $-4n$ may be written as $ax^2 + 2bxy + cy^2$. The basic idea of the proof is to count the number of reduced forms that satisfy $2b = 0, a = 2b$ or

$a = c$, for by Lemma 3.10, this gives the number of classes of order ≤ 2 in $C(-4n)$. Since n is odd, note that r is the number of prime divisors of n.

First, consider forms with $2b = 0$, i.e., the forms $ax^2 + cy^2$, where $ac = n$. Since a and c must be relatively prime and positive, there are 2^r choices for a. To be reduced, we must also have $a < c$, so that we get 2^{r-1} reduced forms of this type.

Next consider forms with $a = 2b$ or $a = c$. Write $n = bk$, where b and k are relatively prime and $0 < b < k$. As above, there are 2^{r-1} such b's. Set $c = (b+k)/2$, and consider the form $2bx^2 + 2bxy + cy^2$. One computes that it has discriminant $-4n$, and since $n \equiv 1 \bmod 4$, its coefficients are relatively prime. We then get 2^{r-1} reduced forms as follows:

$2b < c$: Here, $2bx^2 + 2bxy + cy^2$ is a reduced form.

$2b > c$: Here, $2bx^2 + 2bxy + cy^2$ is properly equivalent to
$cx^2 + 2(c-b)xy + cy^2$ via $(x,y) \mapsto (-y, x+y)$.
Since $2b > c \Rightarrow 2(c-b) < c$, the latter is reduced.

The next step is to check that this process gives *all* reduced forms with $a = 2b$ or $a = c$. We leave this to the reader (see Exercise 3.10).

We thus have $2^{r-1} + 2^{r-1} = 2^r$ elements of order ≤ 2, which shows that $\mu = r + 1$ in this case. The remaining cases are similar and are left to the reader (see Exercise 3.10, Flath [36, §V.5], Gauss [41, §257–258] or Mathews [78, pp. 171–173]). Q.E.D.

This is not the last we will see of the number μ, for it also plays an important role in genus theory.

B. Genus Theory

As in §2, we define two forms of discriminant D to be in the same genus if they represent the same values in $(\mathbb{Z}/D\mathbb{Z})^*$. Let's recall the classification of genera given in §2. Consider the subgroups $H \subset \ker(\chi) \in (\mathbb{Z}/D\mathbb{Z})^*$, where H consists of the values represented by the principal form, and $\chi : (\mathbb{Z}/D\mathbb{Z})^* \longrightarrow \{\pm 1\}$ is defined by $\chi([p]) = (D/p)$ for $p \nmid D$ prime. Then the key result was Lemma 2.24, where we proved that the values represented in $(\mathbb{Z}/D\mathbb{Z})^*$ by a given form $f(x,y)$ are a coset of H in $\ker(\chi)$. This coset determines which genus $f(x,y)$ is in.

Our first step is to relate this theory to the class group $C(D)$. Since all forms in a given class represent the same numbers, sending the class to the coset of $H \subset \ker(\chi)$ it represents defines a map

$$\Phi : C(D) \longrightarrow \ker(\chi)/H. \tag{3.12}$$

Note that a given fiber $\Phi^{-1}(H')$, $H' \in \ker(\chi)/H$, consists of all classes in a given genus (this is what we called the *genus of* H' in Theorem 2.26), and the image of Φ may thus be identified with the set of genera. A crucial observation is that Φ is a group homomorphism:

Lemma 3.13. *The map Φ which maps a class in $C(D)$ to the coset of values represented in $\ker(\chi)/H$ is a group homomorphism.*

Proof. Let $f(x,y)$ and $g(x,y)$ be two forms of discriminant D taking values in the cosets H' and H'' respectively. We can assume that their Dirichlet composition $F(x,y)$ is defined, so that a product of values represented by $f(x,y)$ and $g(x,y)$ is represented by $F(x,y)$. Then $F(x,y)$ represents values in $H'H''$, which proves that $H'H''$ is the coset associated to the composition of $f(x,y)$ and $g(x,y)$. Thus Φ is a homomorphism. Q.E.D.

This lemma has the following consequences:

Corollary 3.14. *Let $D \equiv 0,1 \bmod 4$ be negative. Then:*

(i) *All genera of forms of discriminant D consist of the same number of classes.*

(ii) *The number of genera of forms of discriminant D is a power of two.*

Proof. The first statement follows since all fibers of a homomorphism have the same number of elements. To prove the second, first note that the subgroup H contains all squares in $(\mathbb{Z}/D\mathbb{Z})^*$. This is obvious because if $f(x,y)$ is the principal form, then $f(x,0) = x^2$. Thus every element in $\ker(\chi)/H$ has order ≤ 2, and it follows from the structure theorem for finite Abelian groups that $\ker(\chi)/H \simeq \{\pm 1\}^m$ for some m. Thus the image of Φ, being a subgroup of $\ker(\chi)/H$, has order 2^k for some k. Since $\Phi(C(D))$ tells us the number of genera, we are done. Q.E.D.

Note also that $\Phi(C(D))$ gives a natural group structure on the set of genera, or as Gauss would say, one can define the composition of genera [41, §§246–247].

These elementary facts are nice, but they aren't the whole story. The real depth of the relation between composition and genera is indicated by the following theorem:

Theorem 3.15. *Let $D \equiv 0,1 \bmod 4$ be negative. Then:*

(i) *There are $2^{\mu-1}$ genera of forms of discriminant D, where μ is the number defined in Proposition 3.11.*

(ii) *The principal genus (the genus containing the principal form) consists of the classes in $C(D)^2$, the subgroup of squares in the class group $C(D)$. Thus every form in the principal genus arises by duplication.*

Proof. We first need to give a more efficient method for determining when two forms are in the same genus. The basic idea is to use certain *assigned characters,* which are defined as follows. Let $p_1, \ldots, p_r$ be the distinct odd primes dividing D. Then consider the functions:

$$\chi_i(a) = \left(\frac{a}{p_i}\right) \qquad \text{defined for } a \text{ prime to } p_i,\ i = 1, \ldots, r$$

$$\delta(a) = (-1)^{(a-1)/2} \qquad \text{defined for } a \text{ odd}$$

$$\epsilon(a) = (-1)^{(a^2-1)/8} \qquad \text{defined for } a \text{ odd.}$$

Rather than using all of these functions, we assign only certain ones, depending on the discriminant D. When $D \equiv 1 \bmod 4$, we define $\chi_1, \ldots, \chi_r$ to be the *assigned characters*, and when $D \equiv 0 \bmod 4$, we write $D = -4n$, and then the *assigned characters* are defined by the following table:

n	assigned characters
$n \equiv 3 \bmod 4$	$\chi_1, \cdots, \chi_r$
$n \equiv 1 \bmod 4$	$\chi_1, \cdots, \chi_r, \delta$
$n \equiv 2 \bmod 8$	$\chi_1, \ldots, \chi_r, \delta\epsilon$
$n \equiv 6 \bmod 8$	$\chi_1, \ldots, \chi_r, \epsilon$
$n \equiv 4 \bmod 8$	$\chi_1, \ldots, \chi_r, \delta$
$n \equiv 0 \bmod 8$	$\chi_1, \ldots, \chi_r, \delta, \epsilon$

Note that the number of assigned characters is exactly the number μ given in Proposition 3.11. It is easy to see that the assigned characters give a homomorphism

$$\Psi : (\mathbb{Z}/D\mathbb{Z})^* \longrightarrow \{\pm 1\}^\mu. \tag{3.16}$$

The crucial property of Ψ is the following:

Lemma 3.17. *The homomorphism $\Psi : (\mathbb{Z}/D\mathbb{Z})^* \longrightarrow \{\pm 1\}^\mu$ of (3.16) is surjective and its kernel is the subgroup H of values represented by the principal form. Thus Ψ induces an isomorphism*

$$(\mathbb{Z}/D\mathbb{Z})^*/H \xrightarrow{\sim} \{\pm 1\}^\mu.$$

Proof. When $D \equiv 1 \bmod 4$, the proof is quite easy. First note that if p is an odd prime, then for any $m \geq 1$, the Legendre symbol (a/p) induces a surjective homomorphism

$$(\cdot/p) : (\mathbb{Z}/p^m\mathbb{Z})^* \longrightarrow \{\pm 1\} \tag{3.18}$$

whose kernel is exactly the subgroup of squares of $(\mathbb{Z}/p^m\mathbb{Z})^*$ (see Exercise 3.11). Now let $D = -\prod_{i=1}^{\mu} p_i^{m_i}$ be the prime factorization of D. The Chinese Remainder Theorem tells us that

$$(\mathbb{Z}/D\mathbb{Z})^* \xrightarrow{\sim} \prod_{i=1}^{\mu} (\mathbb{Z}/p_i^{m_i}\mathbb{Z})^*,$$

so that the map Ψ can be interpreted as the map

$$\prod_{i=1}^{\mu} (\mathbb{Z}/p_i^{m_i}\mathbb{Z})^* \longrightarrow \{\pm 1\}^\mu$$

given by $([a_1], \ldots, [a_\mu]) \mapsto ((a_1/p_1), \ldots, (a_\mu/p_\mu))$. By the analysis of (3.18), it follows that Ψ is surjective and its kernel is exactly the subgroup of squares of $(\mathbb{Z}/D\mathbb{Z})^*$. By part (c) of Exercise 2.17, this equals the subgroup H of values represented by the principal form $x^2 + xy + ((1-D)/4)y^2$, and we are done.

The proof is more complicated when $D = -4n$, mainly because the subgroup H represented by $x^2 + ny^2$ may be slightly larger than the subgroup of squares. However, the above argument using the Chinese Remainder Theorem can be adapted to

this case. The odd primes dividing n are no problem, but 2 causes considerable difficulty (see Exercise 3.11 for the details). Q.E.D.

We can now prove Theorem 3.15. To prove (i), note that $\ker(\chi)$ has index 2 in $(\mathbb{Z}/D\mathbb{Z})^*$. By Lemma 3.17, it follows that $\ker(\chi)/H$ has order $2^{\mu-1}$. We know that the number of genera is the order of $\Phi(C(D)) \subset \ker(\chi)/H$, so that it suffices to show $\Phi(C(D)) = \ker(\chi)/H$. Since Φ maps a class to the coset of values it represents, we need to show that every congruence class in $\ker(\chi)$ contains a number represented by a form of discriminant D. This is easy: Dirichlet's theorem on primes in arithmetic progressions tells us that any class in $\ker(\chi)$ contains an odd prime p. But $[p] \in \ker(\chi)$ means that $\chi([p]) = (D/p) = 1$, so that by Lemma 2.5, p is represented by a form of discriminant D, and (i) is proved.

To prove (ii), let C denote the class group $C(D)$. Since $\Phi : C \to \ker(\chi)/H \simeq \{\pm 1\}^{\mu-1}$ is a homomorphism, it follows that $C^2 \subset \ker(\Phi)$, and we get an induced map

$$C/C^2 \longrightarrow \{\pm 1\}^{\mu-1}. \tag{3.19}$$

We compute the order of C/C^2 as follows. The squaring map from C to itself gives a short exact sequence

$$0 \to C_0 \to C \to C^2 \to 0$$

where C_0 is the subgroup of C of elements of order ≤ 2. It follows that the index $[C : C^2]$ equals the order of C_0, which is $2^{\mu-1}$ by Proposition 3.11.

Thus, in the map given in (3.19), both the domain and the range have the same order. But from (i) we know that the map is surjective, so that it must be an isomorphism. Hence C^2 is exactly the kernel of the map Φ. Since $\ker(\Phi)$ consists of the classes in the principal genus, the theorem is proved. Q.E.D.

We have now proved the main theorems of genus theory for primitive positive definite forms. These results are due to Gauss and appear in the fifth section of *Disquisitiones Arithmeticae* [41, §§229–287]. Gauss' treatment is more general than ours, for he considers both the definite and indefinite forms, and in particular, he shows that Proposition 3.11 and Theorem 3.15 are true for *any* nonsquare discriminant, positive or negative. His proofs are quite difficult, and at the end of this long series of arguments, Gauss makes the following comment about genus theory [41, §287]:

> these theorems are among the most beautiful in the theory of binary forms, especially because, despite their extreme simplicity, they are so profound that a rigorous demonstration requires the help of many other investigations.

Besides these theorems, there is another component to Gauss' genus theory not mentioned so far: Gauss' second proof of quadratic reciprocity [41, §262], which uses the genus theory developed above. We will not discuss Gauss' proof since it uses forms of positive discriminant, though the main ideas of the proof are outlined in Exercises 3.12 and 3.13. Many people regard this as the deepest of Gauss' many proofs of quadratic reciprocity.

Gauss' approach to genus theory is somewhat different from ours. In *Disquisitiones,* genera are defined in terms of the *assigned characters* introduced in the proof of Theorem 3.15. Given a form $f(x,y)$ of discriminant D, let $f(x,y)$ represent a number a relatively prime to D. If the μ assigned characters are evaluated at a, then Gauss calls the resulting μ-tuple the *complete character* of $f(x,y)$, and he defines two forms of discriminant D to be in the same genus if they have the same complete character [41, §231]. The following lemma shows that this is equivalent to our previous definition of genus:

Lemma 3.20. *The complete character depends only on the form $f(x,y)$, and two forms of discriminant D lie in the same genus (as defined in §2) if and only if they have the same complete character.*

Proof. Suppose that $f(x,y)$ represents a, where a is relatively prime to D. Then Gauss' complete character is nothing other than $\Psi([a])$, where Ψ is the map defined in (3.16). By Lemma 2.24, the possible a's lie in a coset H' of H in $(\mathbb{Z}/D\mathbb{Z})^*$, and this coset determines the genus of $f(x,y)$. Using Lemma 3.17, it follows that the complete character is uniquely determined by H', and Lemma 3.20 is proved. Q.E.D.

We should mention that Gauss' use of the word "character" is where the modern term "group character" comes from. Also, it is interesting to note that Gauss never mentions the connection between his characters and Lagrange's implicit genus theory. While Gauss' characters make it easy to decide when two forms belong to the same genus (see Exercise 3.14 for an example), they are not very intuitive. Unfortunately, most of Gauss' successors followed his presentation of genus theory, so that readers were presented with long lists of characters and no motivation whatsoever. The simple idea of grouping forms according to the congruence classes they represent was usually not mentioned. This happens in Dirichlet [28, pp. 313–316] and in Mathews [78, pp. 132–136], although Smith [95, pp. 202–207] does discuss congruence classes.

So far we have discussed two ways to formulate genera, Lagrange's and Gauss'. There are many other ways to state the definition, but before we can discuss them, we need some terminology. We say that two forms $f(x,y)$ and $g(x,y)$ are *equivalent over a ring R* if there is a matrix $\left(\begin{smallmatrix} p & q \\ r & s \end{smallmatrix}\right) \in \mathrm{GL}(2,R)$ such that $f(x,y) = g(px+qy, rx+sy)$. If $R = \mathbb{Z}/m\mathbb{Z}$, we say that $f(x,y)$ and $g(x,y)$ are *equivalent modulo m.* We then have the following theorem:

Theorem 3.21. *Let $f(x,y)$ and $g(x,y)$ be primitive forms of discriminant $D \neq 0$, positive definite if $D < 0$. Then the following statements are equivalent:*

(i) *$f(x,y)$ and $g(x,y)$ are in the same genus, i.e., they represent the same values in $(\mathbb{Z}/D\mathbb{Z})^*$.*

(ii) *$f(x,y)$ and $g(x,y)$ represent the same values in $(\mathbb{Z}/m\mathbb{Z})^*$ for all nonzero integers m.*

(iii) *$f(x,y)$ and $g(x,y)$ are equivalent modulo m for all nonzero integers m.*

(iv) *$f(x,y)$ and $g(x,y)$ are equivalent over the p-adic integers $\mathbb{Z}_p$ for all primes p.*

(v) *$f(x,y)$ and $g(x,y)$ are equivalent over $\mathbb{Q}$ via a matrix in $\mathrm{GL}(2,\mathbb{Q})$ whose entries have denominators prime to $2D$.*

(vi) *$f(x,y)$ and $g(x,y)$ are equivalent over $\mathbb{Q}$ without essential denominator, i.e., given any nonzero m, a matrix in $\mathrm{GL}(2,\mathbb{Q})$ can be found which takes one form to the other and whose entries have denominators prime to m.*

Proof. It is easy to prove (vi) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) and (vi) $\Rightarrow$ (v) $\Rightarrow$ (i) (see Exercise 3.15), and (iii) $\Leftrightarrow$ (iv) is a standard argument using the compactness of $\mathbb{Z}_p$ (see Borevich and Shafarevich [8, p. 41] for an analogous case). A proof of (i) $\Rightarrow$ (iii) appears in Hua [57, §12.5, Exercise 4], and (i) $\Rightarrow$ (iv) is in Jones [63, pp. 103–104]. Finally, the implication (iv) $\Rightarrow$ (vi) uses the Hasse principle for the equivalence of forms over $\mathbb{Q}$ and may be found in Jones [63, Theorem 40] or Siegel [91]. Q.E.D.

Some modern texts give yet a different definition, saying that two forms are in the same genus if and only if they are equivalent over $\mathbb{Q}$ (see, for example, Borevich and Shafarevich [8, p. 241]). This characterization doesn't hold in general ($x^2 + 18y^2$ and $2x^2 + 9y^2$ are rationally equivalent but belong to different genera—see Exercise 3.16), but it does work for *field discriminants,* which means that $D \equiv 1 \bmod 4$, D squarefree, or $D = 4k$, $k \not\equiv 1 \bmod 4$, k squarefree (see Exercise 3.17—we will study such discriminants in more detail in §5). According to Dickson [26, Vol. III, pp. 216 and 236], Eisenstein suggested in 1852 that genera could be defined using rational equivalence, and only later, in 1867, did Smith point out that extra assumptions are needed on the denominators.

C. $p = x^2 + ny^2$ and Euler's Convenient Numbers

Our discussion of genus theory has distracted us from our problem of determining when a prime p can be written as $x^2 + ny^2$. Recall from Corollary 2.27 that genus theory gives us congruence conditions for p to be represented by a reduced form in the principal genus. The nicest case is when every genus of discriminant $-4n$ consists of a single class, for then we get congruence conditions that characterize $p = x^2 + ny^2$ (this is what made the examples in (2.28) work). Let's see if the genus theory developed in this section can shed any light on this special case. We have the following result:

Theorem 3.22. *Let n be a positive integer. Then the following statements are equivalent:*

(i) *Every genus of forms of discriminant $-4n$ consists of a single class.*

(ii) *If $ax^2 + bxy + cy^2$ is a reduced form of discriminant $-4n$, then either $b = 0$, $a = b$ or $a = c$.*

(iii) *Two forms of discriminant $-4n$ are equivalent if and only if they are property equivalent.*

(iv) *The class group $C(-4n)$ is isomorphic to $(\mathbb{Z}/2\mathbb{Z})^m$ for some integer m.*

(v) *The class number $h(-4n)$ equals $2^{\mu-1}$, where μ is as in Proposition* 3.11.

Proof. We will prove (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (v) $\Rightarrow$ (i). Let C denote the class group $C(-4n)$.

Since the principal genus is C^2 by Theorem 3.15, (i) implies that $C^2 = \{1\}$, so that every element of C has order ≤ 2. Then Lemma 3.10 shows that (i) $\Rightarrow$ (ii).

Next assume (ii), and suppose that two forms of discriminant $-4n$ are equivalent. By Exercise 3.8, we know that one is properly equivalent to the other or its opposite. We may assume that the forms are reduced, so that by assumption $b = 0$, $a = b$ or $a = c$. The proof of Theorem 2.8 shows that forms of this type are always properly equivalent to their opposites, so that the forms are properly equivalent. This proves (ii) $\Rightarrow$ (iii).

Recall that any form is equivalent to its opposite via $(x,y) \mapsto (x,-y)$. Thus (iii) implies that any form and its opposite lie in the same class in C. Since the opposite gives the inverse in C by Theorem 3.9, we see that every class is its own inverse. The structure theorem for finite Abelian groups shows that the only groups with this property are $(\mathbb{Z}/2\mathbb{Z})^m$, and (iii) $\Rightarrow$ (iv) is proved.

Next, Theorem 3.15 implies that the number of genera is $[C : C^2] = 2^{\mu-1}$, so that

$$h(-4n) = |C| = [C : C^2]|C^2| = 2^{\mu-1}|C^2|. \tag{3.23}$$

If (iv) holds, then $C^2 = \{1\}$, and then (v) follows immediately from (3.23). Finally, given (v), (3.23) implies that $C^2 = \{1\}$, so that by Theorem 3.15, the principal genus consists of a single class. Since every genus consists of the same number of classes, (i) follows, and the theorem is proved. Q.E.D.

Notice how this theorem runs the full gamut of what we've done so far: the conditions of Theorem 3.22 involve genera, reduced forms, the class number, the structure of the class group and the relation between equivalence and proper equivalence. For computational purposes, the last condition (v) is especially useful, for it only requires knowing the class number. This makes it much easier to verify that the examples in (2.28) have only one class per genus.

Near the end of the fifth section of *Disquisitiones,* Gauss lists 65 discriminants that satisfy this theorem [41, §303]. Grouped according to class number, they are:

$h(-4n)$	n's with one class per genus
1	1,2,3,4,7
2	5,6,8,9,10,12,13,15,16,18,22,25,28,37,58
4	21,24,30,33,40,42,45,48,57,60,70,72,78,85,88,93,102,112 130,133,177,190,232,253
8	105,120,165,168,210,240,273,280,312,330,345,357,385 408,462,520,760
16	840,1320,1365,1848

Gauss was interested in these 65 n's not for their relation to the question of when $p = x^2 + ny^2$, but rather because they had been discovered earlier by Euler in a different context. Euler called a number n a *convenient number* (numerus idoneus) if it satisfies the following criterion:

> Let m be an odd number relatively prime to n which is properly represented by $x^2 + ny^2$. If the equation $m = x^2 + ny^2$ has only one solution with $x, y \geq 0$, then m is a prime number.

Euler was interested in convenient numbers because they helped him find large primes. For example, working with $n = 1848$, he was able to show that

$$18,518,809 = 197^2 + 1848 \cdot 100^2$$

is prime, a large one for Euler's time. Convenient numbers are a fascinating topic, and the reader should consult Frei [38] or Weil [106, pp. 219–226] for a fuller discussion. We will confine ourselves to the following remarkable observation of Gauss:

Proposition 3.24. *A positive integer n is a convenient number if and only if for forms of discriminant $-4n$, every genus consists of a single class.*

Proof. We begin with a lemma:

Lemma 3.25. *Let m be a positive odd number relatively prime to $n > 1$. Then the number of ways that m is properly represented by a reduced form of discriminant $-4n$ is*

$$2\prod_{p|m}\left(1 + \left(\frac{-n}{p}\right)\right).$$

Proof. See Exercise 3.20 or Landau [71, Vol. 1, p. 144]. Q.E.D.

This classical lemma belongs to an area of quadratic forms that we have ignored, namely the study of the *number* of representations of a number by a form. To see what this has to do with genus theory, note that two forms representing m must lie in the same genus, for the values they represent in $(\mathbb{Z}/4n\mathbb{Z})^*$ are not disjoint. We thus get the following corollary of Lemma 3.25:

Corollary 3.26. *Let m be properly represented by a primitive positive definite form $f(x,y)$ of discriminant $-4n$, $n > 1$, and assume that m is odd and relatively prime to n. If r denotes the number of prime divisors of m, then m is properly represented in exactly 2^{r+1} ways by a reduced form in the genus of $f(x,y)$.* Q.E.D.

Now we can prove the proposition. First, assume that there is only one class per genus. If m is properly represented by $x^2 + ny^2$ and $m = x^2 + ny^2$ has a unique solution when $x, y \geq 0$, then we need to prove that m is prime. The above corollary shows that m is properly represented by $x^2 + ny^2$ in 2^{r+1} ways since $x^2 + ny^2$ is the only reduced form in its genus. At least 2^{r-1} of these representations satisfy $x, y \geq 0$, and then our assumption on m implies that $r = 1$, i.e., m is a prime power p^a. If $a \geq 2$, then Lemma 3.25 shows that p^{a-2} also has a proper representation, and it follows easily

that m has at least two representations in nonnegative integers. This contradiction proves that m is prime, and hence n is a convenient number.

Conversely, assume that n is convenient. Let $f(x,y)$ be a form of discriminant $-4n$, and let $g(x,y)$ be the composition of $f(x,y)$ with itself. We can assume that $g(x,y)$ is reduced, and it suffices to show that $g(x,y) = x^2 + ny^2$ (for then every element in the class group has order ≤ 2, which by Theorem 3.22 implies that there is one class per genus).

Assume that $g(x,y) \neq x^2 + ny^2$, and let p and q be distinct odd primes not dividing n which are represented by $f(x,y)$. (In §9 we will prove that $f(x,y)$ represents infinitely many primes.) Then $g(x,y)$ represents pq, and formula (2.31) shows that $x^2 + ny^2$ does too. By Corollary 3.26, pq has only 8 proper representations by reduced forms of discriminant $-4n$. At least one comes from $g(x,y)$, leaving at most 7 for $x^2 + ny^2$. It follows that pq is uniquely represented by $x^2 + ny^2$ when we restrict to nonnegative integers. This contradicts our assumption that n is convenient. Q.E.D.

Gauss never states Proposition 3.24 formally, but it is implicit in the methods he discusses for factoring large numbers [41, §§329–334].

In §2 we asked how many such n's there were. Gauss suggests [41, §303] that the 65 given by Euler are the only ones. In 1934 Chowla [17] proved that the number of such n's is finite, and by 1973 it was known that Euler's list is complete except for possibly one more n (see Weinberger [108]). Whether or not this last n actually exists is still an open question.

From our point of view, the upshot is that there are only finitely many theorems like (2.28) where $p = x^2 + ny^2$ is characterized by simple congruences modulo $4n$. Thus genus theory cannot solve our basic question for all n. In some cases, such as $D = -108$, it's completely useless (all three reduced forms $x^2 + 27y^2$ and $4x^2 \pm 2xy + 7y^2$ lie in the same genus), and even when it's a partial help, such as $D = -56$, we're still stuck (we can separate $x^2 + 14y^2$ and $2x^2 + 7y^2$ from $3x^2 \pm 2xy + 5y^2$, but we can't distinguish between the first two). And notice that by part (iii) of Theorem 3.21, forms in the same genus are equivalent modulo m for *all* $m \neq 0$, so that no matter how m is chosen, there are no congruences $p \equiv a, b, c, \ldots \bmod m$ which can separate forms in the same genus. Something new is needed. In 1833, Dirichlet described the situation as follows [27, Vol. I, p. 201]:

> there lies in the mentioned [genus] theory an incompleteness, in that it certainly shows that a prime number, as soon as it is contained in a linear form [congruence class], necessarily must assume one of the corresponding quadratic forms, only without giving any a priori method for deciding which quadratic form it will be. ... It becomes clear that the characteristic property of a single quadratic form belonging to a group [genus] cannot be expressed through the prime numbers in the corresponding linear forms, but necessarily must be expressed by another theory not depending on the elements at hand.

As we already know from Euler's conjectures concerning $x^2 + 27y^2$ and $x^2 + 64y^2$ (see (1.22) and (1.23)), the new theory we're seeking involves residues of higher powers. Gauss rediscovered Euler's conjectures in 1805, and he proved them in the course of his work on cubic and biquadratic reciprocity. In §4 we will give careful

statements of these reciprocity theorems and show how they can be used to prove Euler's conjectures.

D. Disquisitiones Arithmeticae

Gauss' *Disquisitiones Arithmeticae* covers a wide range of topics in number theory, including congruences, quadratic reciprocity, quadratic forms (in two and three variables), and the cyclotomic fields $\mathbb{Q}(\zeta_n)$, $\zeta_n = e^{2\pi i/n}$. There are several excellent accounts of what's in *Disquisitiones,* notably Bühler [13, Chapter 3], Bachmann [42, Vol. X.2.1, pp. 8–40] and Rieger [84], and translations into English and German are available (see item [41] in the references). Rather than try to survey the whole book, we will instead make some comments on Gauss' treatment of quadratic reciprocity and quadratic forms, for in each case he does things slightly differently from the theory presented in §§2 and 3.

Disquisitiones contains the first published (valid) proof of the law of quadratic reciprocity. One surprise is that Gauss never uses the term "quadratic reciprocity." Instead, Gauss uses the phrase "fundamental theorem," which he explains as follows [41, §131]:

> Since almost everything that can be said about quadratic residues depends on this theorem, the term *fundamental theorem* which we will use from now on should be acceptable.

In the more informal setting of his mathematical diary, Gauss uses the term "golden theorem" to describe his high regard for quadratic reciprocity [42, Vol. X.1, entries 16, 23 and 30 on pp. 496–501] (see Gray [44] for an English translation). Likewise absent from *Disquisitiones* is the Legendre symbol, for Gauss uses the notation *aRb* or *aNb* to indicate whether or not a was a quadratic residue modulo b [41, §131]. (The Legendre symbol does appear in some of his handwritten notes—see [42, Vol. X.1, p. 53]—but this doesn't happen very often.)

One reason why Gauss ignored Legendre's terminology is that Gauss discovered quadratic reciprocity independently of his predecessors. In a marginal note in his copy of *Disquisitiones,* Gauss states that "we discovered the fundamental theorem by induction in March 1795. We found our first proof, the one contained in this section, April 1796" [41, p. 468, English editions] or [42, Vol. I, p. 476]. In 1795 Gauss was still a student at the Collegium Carolinum in Brunswick, and only later, while at Göttingen, did he discover the earlier work of Euler and Legendre on reciprocity.

Gauss' proof from April 1796 appears in §§135–144 of *Disquisitiones.* The theorem is stated in two forms: the usual version of quadratic reciprocity appears in [41, §131], and the more general version that holds for the Jacobi symbol (which we used in the proof of Lemma 1.14) is given in [41, §133]. The proof uses complete induction on the prime p, and there are many cases to consider, some of which use reciprocity for the Jacobi symbol (which would hold for numbers smaller than p). As Gauss wrote in 1808, the proof "proceeds by laborious steps and is burdened by detailed calculations" [42, Vol. II, p. 4]. In 1857, Dirichlet used the Jacobi symbol to simplify the proof and reduce the number of cases to just two [27, Vol. II, pp. 121–138]. It is interesting to note that what Gauss proves in *Disquisitiones* is actually a bit

more general than the usual statement of quadratic reciprocity for the Jacobi symbol (see Exercise 3.24). Thus, when Jacobi introduced the Jacobi symbol in 1837 [61, Vol. VI, p. 262], he was simply giving a nicer but less general formulation of what was already in *Disquisitiones*.

As we mentioned in our discussion of genus theory, *Disquisitiones* also contains a second proof of reciprocity that is quite different in nature. The first proof is awkward but elementary, while the second uses Gauss' genus theory and is much more sophisticated.

Gauss' treatment of quadratic forms occupies the fifth (and longest) section of *Disquisitiones*. It is not easy reading, for many of the arguments are very complicated. Fortunately, there are more modern texts that cover pretty much the same material (in particular, see either Flath [36] or Mathews [78]). Gauss starts with the case of positive definite forms, and the theory he develops is similar to the first part of §2. Then, in [41, §182], he gives some applications to number theory, which are introduced as follows:

> Let us now consider certain particular cases both because of their remarkable elegance and because of the painstaking work done on them by Euler, who endowed them with an almost classical distinction.

As might be expected, Gauss first proves Fermat's three theorems (1.1), and then he proves Euler's conjecture for $p = x^2 + 5y^2$ using Lagrange's implicit genus theory (his proof is similar to what we did in (2.19), (2.20) and (2.22)). Interestingly enough, Gauss never mentions the relation between this example and genus theory. In contrast to Lagrange and Legendre, Gauss works out few examples. His one comment is that "the reader can derive this proposition [concerning $x^2 + 5y^2$] and an infinite number of other particular ones from the preceding and the following discussions" [41, §182].

Gauss always assumed that the middle coefficient was even, so that his forms were written $f(x,y) = ax^2 + 2bxy + cy^2$. He used the ordered triple (a,b,c) to denote $f(x,y)$ [41, §153], and he defined its *determinant* to be $b^2 - ac$ [41, §154]. Note that the discriminant of $ax^2 + 2bxy + cy^2$ is just 4 times Gauss' determinant.

Gauss did not assume that the coefficients of his forms were relatively prime, and he organized forms into *orders* according to the common divisors of the coefficients. More precisely, the forms $ax^2 + 2bxy + cy^2$ and $a'x^2 + 2b'xy + c'y^2$ are in the same *order* provided that $\gcd(a,b,c) = \gcd(a',b',c')$ and $\gcd(a,2b,c) = \gcd(a',2b',c')$ [41, §226]. To get a better idea of how this works, consider a primitive quadratic form $ax^2 + bxy + cy^2$. Here, a,b and c are relatively prime integers, and b may be even or odd. We can fit this form into Gauss' scheme as follows:

b even: Then $b = 2b'$, and $ax^2 + 2b'xy + cy^2$ satisfies $\gcd(a,b',c) = \gcd(a,2b',c) = 1$. Gauss called forms in this order *properly primitive.*

b odd: Then $2ax^2 + 2bxy + 2cy^2$ satisfies $\gcd(2a,b,2c) = 1$ and $\gcd(2a,2b,2c) = 2$. He called forms in this order *improperly primitive.*

All primitive forms are present, though the ones with b odd appear in disguised form. This doesn't affect the class number but does cause problems with composition.

Gauss' classification of forms thus consists of orders, which are made up of genera, which are in turn made up of classes. This is reminiscent of the Linnean classification in biology, where the categories are class, order, family, genus and species. Gauss' terms all appear on Linneaus' list, and it is thus likely that this is where Gauss got his terminology. Since our current term "equivalence class" comes from Gauss' example of *classes* of properly *equivalent* forms, we see that there is an unexpected link between modern set theory and eighteenth-century biology.

Finally, let's make one comment about composition. Gauss' theory of composition has always been one of the more difficult parts of *Disquisitiones* to read, and part of the reason is the complexity of Gauss' presentation. For example, the proof that composition is associative involves checking that 28 equations are satisfied [41, §240]. But a multiplicity of equations is not the only difficulty here—there is also an interesting conceptual issue. Namely, in order to define the class group, notice that Gauss has to put the structure of an abstract Abelian group on a set of equivalence classes. Considering that we're talking about the year 1801, this is an amazing level of abstraction. But then, *Disquisitiones* is an amazing book.

E. Exercises

3.1. Assume that $F(x,y) = Ax^2 + Bxy + Cy^2$ is the composition of the two forms $f(x,y) = ax^2 + bxy + cy^2$ and $g(x,y) = a'x^2 + b'xy + c'y^2$ via

$$\begin{aligned} f(x,y)g(z,w) = F(&a_1xz + b_1xw + c_1yz + d_1yw, a_2xz \\ &+ b_2xw + c_2yz + d_2yw), \end{aligned}$$

and suppose that all three forms have discriminant $D \neq 0$. The goal of this exercise is to prove Gauss' formulas (3.1).

(a) By specializing the variables x, y, z and w, prove that

$$\begin{aligned} aa' &= Aa_1^2 + Ba_1a_2 + Ca_2^2 \\ ac' &= Ab_1^2 + Bb_1b_2 + Cb_2^2 \\ ab' &= 2Aa_1b_1 + B(a_1b_2 + a_2b_1) + 2Ca_2b_2. \end{aligned}$$

Hint: for the first one, try $x = z = 1$ and $y = w = 0$.

(b) Prove that $a = \pm(a_1b_2 - a_2b_1)$. Hint: prove that

$$a^2(b'^2 - 4a'c') = (a_1b_2 - a_2b_1)^2(B^2 - 4AC).$$

(c) Prove that $a' = \pm(a_1c_2 - a_2c_1)$.

3.2. Show that the compositions given in (2.30) and (2.31) are not direct compositions.

3.3. Prove Lemma 3.5. Hint: there are $a, a_1, \ldots, a_r$ such that $am + \Sigma_{i=1}^r a_i p_i = 1$.

3.4. Verify that the congruences (3.4) satisfy the compatibility conditions stated in Lemma 3.5.

3.5. Assume that $f(x,y) = ax^2+bxy+cy^2$, $g(x,y) = a'x^2+b'xy+c'y^2$ and B are as in Lemma 3.2. We want to show that $aa'x^2+Bxy+Cy^2$, $C = (B^2-D)/4aa'$, is the direct composition of $f(x,y)$ and $g(x,y)$.

(a) Show that $f(x,y)$ and $g(x,y)$ are properly equivalent to $ax^2+Bxy+a'Cy^2$ and $a'x^2+Bxy+aCy^2$ respectively. Hint: use $B \equiv b \bmod 2a$ for $f(x,y)$.

(b) Let $X = xz-Cyw$ and $Y = axw+a'yz+Byw$. Then show that

$$\begin{aligned}(ax^2+Bxy+a'Cy^2)(a'z^2+Bzw+aCw^2)\\ = aa'X^2+BXY+CY^2.\end{aligned}$$

Furthermore, show that this is a direct composition in the sense of (3.1). Hint: first show that

$$\begin{aligned}(ax+(B+\sqrt{D})y/2)(a'z+(B+\sqrt{D})w/2)\\ = aa'X+(B+\sqrt{D})Y/2.\end{aligned}$$

(c) Suppose that a form $G(x,y)$ is the direct composition of forms $h(x,y)$ and $k(x,y)$. If $\tilde{h}(x,y)$ is properly equivalent to $h(x,y)$, then show that $G(x,y)$ is also the direct composition of $\tilde{h}(x,y)$ and $k(x,y)$.

(d) Use (a)–(c) to show that Dirichlet composition is a direct composition.

3.6. This problem studies the relation between Legendre's and Dirichlet's formulas for composition.

(a) Suppose that $f(x,y) = ax^2+2bxy+cy^2$ and $g(x,y) = a'x^2+2b'xy+c'y^2$ have the same discriminant and satisfy $\gcd(a,a') = 1$. Show that the Dirichlet composition of these forms is the one given by Legendre's formula with both signs + in (2.32).

(b) In Exercise 2.26, we saw that $14x^2+10xy+21y^2$ and $9x^2+2xy+30y^2$ can be composed to obtain $126x^2 \pm 74xy+13y^2$ and $126x^2 \pm 38xy+5y^2$. Which one of these four is the direct composition of the original two forms?

3.7. Show that $acx^2+bxy+y^2$ is properly equivalent to the principal form.

3.8. For us, a *class* consists of all forms properly equivalent to a given form. Let a *Lagrangian class* (this terminology is due to Weil [106, p. 319]) consist of all forms equivalent (properly or improperly) to a given form.

(a) Prove that the Lagrangian class of a form is the union of the class of the form and the class of its opposite.

(b) Show that the following statements are equivalent:

(i) The Lagrangian class of $f(x,y)$ equals the class of $f(x,y)$.

(ii) $f(x,y)$ is properly equivalent to its opposite.

(iii) $f(x,y)$ is properly and improperly equivalent to itself.

(iv) The class of $f(x,y)$ has order ≤ 2 in the class group.

3.9. In this problem we will describe the "almost" group structure given by Legendre's theory of composition. Let G be an Abelian group and let $\sim$ be the equivalence relation which identifies a^{-1} and a for all $a \in G$.

(a) Show that multiplication on G induces an operation on $G/\sim$ which takes either one or two values. Furthermore, if $a, b \in G$ and $[a]$, $[b]$ are their classes in $G/\sim$, then show that $[a]\cdot[b]$ takes on only one value if and only if a, b or ab has order ≤ 2 in G.

(b) If G is cyclic of order 8, show that $G/\sim$ is isomorphic (in the obvious sense) to the structure given by (2.33) and (2.34).

(c) If $C(D)$ is the class group of forms of discriminant D, show that $C(D)/\sim$ can be naturally identified with the set of Lagrangian classes of forms of discriminant D (see Exercise 3.8).

3.10. Complete the proof of Proposition 3.11 for the case $D = -4n$, $n \equiv 1 \bmod 4$, and prove all of the remaining cases.

3.11. This exercise is concerned with the proof of Lemma 3.17.

(a) Prove that the map (3.18) is surjective and its kernel is the subgroup of squares.

(b) We next want to prove the lemma when $D = -4n, n > 0$. Write $n = 2^a m$ where m is odd, so that we have an isomorphism

$$(\mathbb{Z}/D\mathbb{Z})^* \simeq (\mathbb{Z}/2^{a+2}\mathbb{Z})^* \times (\mathbb{Z}/m\mathbb{Z})^*.$$

Let H denote the subgroup of values represented by x^2+ny^2.

(i) Show that $H = H_1 \times (\mathbb{Z}/m\mathbb{Z})^{*2}$ for $H_1 = H \cap ((\mathbb{Z}/2^{a+2}\mathbb{Z})^* \times \{1\})$.

(ii) When $a \geq 4$, show that $H_1 = (\mathbb{Z}/2^{a+2}\mathbb{Z})^{*2}$, where H_1 is as in (i). Hint: the description of $(\mathbb{Z}/2^{a+2}\mathbb{Z})^*$ given in Ireland and Rosen [59, §4.1] will be useful.

(iii) Prove Lemma 3.17 when $D \equiv 0 \bmod 4$. Hint: treat the cases $a = 0$, 1, 2, 3 and ≥ 4 separately. See also Ireland and Rosen [59, §4.1].

3.12. In Exercises 3.12 and 3.13 we will sketch Gauss' second proof of quadratic reciprocity. There are two parts to the proof: first, one shows, without using quadratic reciprocity, that for any nonsquare discriminant D,

(*) the number of genera of forms of discriminant D is $\leq 2^{\mu-1}$,

where μ is defined in Proposition 3.11, and second, one shows that $(*)$ implies quadratic reciprocity. This exercise will do the first step, and Exercise 3.13 will take care of the second.

We proved in Exercise 2.10 that when $D > 0$ is not a perfect square, there are only finitely many proper equivalence classes of primitive forms of discriminant D. The set of equivalence classes will be denoted $C(D)$, and as in the positive definite case, $C(D)$ becomes a finite Abelian group under Dirichlet composition (we will prove this in the exercises to §7). We will assume that Proposition 3.11 and Theorem 3.15 hold for all nonsquare discriminants D. This is where we pay the price for restricting ourselves to positive definite forms—the proofs in the text only work for $D < 0$. For proofs of these theorems when $D > 0$, see Flath [36, Chapter V], Gauss [41, §§257–258] or Mathews [78, pp. 171–173].

To prove $(*)$, let D be any nonsquare discriminant, and let C denote the class group $C(D)$. Let $H \subset (\mathbb{Z}/D\mathbb{Z})^*$ be the subgroup of values represented by the principal form.

(a) Show that genera can be classified by cosets of H in $(\mathbb{Z}/D\mathbb{Z})^*$. Thus, instead of the map Φ of (3.12), we can use the map

$$\Phi' : C \longrightarrow (\mathbb{Z}/D\mathbb{Z})^*/H,$$

so that $\ker(\Phi')$ is the principal genus and $\Phi'(C)$ is the set of genera. Note that this argument does not use quadratic reciprocity.

(b) Since H contains all squares in $(\mathbb{Z}/D\mathbb{Z})^*$, it follows that $C^2 \subset \ker(\Phi)'$. Now adapt the proof of Theorem 3.15 to show that

$$\text{the number of genera is } \leq [C : C^2] = 2^{\mu-1},$$

where the last equality follows from Proposition 3.11. This proves $(*)$.

3.13. In this exercise we will show that quadratic reciprocity follows from statement $(*)$ of Exercise 3.12. As we saw in §1, it suffices to show

$$\left(\frac{p^*}{q}\right) = 1 \iff \left(\frac{q}{p}\right) = 1,$$

where p and q are distinct odd primes and $p^* = (-1)^{(p-1)/2}p$.

(a) Show that Lemma 3.17 holds for all nonsquare discriminants D, so that we can use the assigned characters to distinguish genera.

(b) Assume that $(p^*/q) = 1$. Applying Lemma 2.5 with $D = p^*$ shows that q is represented by a form $f(x,y)$ of discriminant p^*. The number μ from Proposition 3.11 is 1, so that by $(*)$, there is only one genus. Hence the assigned character (there is only one in this case) must equal 1 on

any number represented by $f(x,y)$, in particular q. Use this to prove that $(q/p) = 1$. This proves that $(p^*/q) = 1 \Rightarrow (q/p) = 1$.

(c) Next, assume that $(q/p) = 1$ and that either $p \equiv 1 \bmod 4$ or $q \equiv 1 \bmod 4$. Use part (b) to show that $(p^*/q) = 1$.

(d) Finally, assume that $(q/p) = 1$ and that $p \equiv q \equiv 3 \bmod 4$. This time we will consider forms of discriminant pq. Proposition 3.11 shows that $\mu = 2$, so that by $(*)$, there are at most two genera. Furthermore, the assigned characters are $\chi_1(a) = (a/p)$ and $\chi_2(a) = (a/q)$. Now consider the form $f(x,y) = px^2 + pxy + ((p-q)/4)y^2$, which is easily seen to have discriminant pq. Letting $(x,y) = (0,2)$, it represents $p-q$. Use this to compute the complete character of the forms $f(x,y)$ and $-f(x,y)$, and show that one of these must lie in the principal genus since there are at most two genera. Then show that $(-p/q) = 1$. Note that parts (c) and (d) imply that $(q/p) = 1 \Rightarrow (p^*/q) = 1$, which completes the proof of quadratic reciprocity.

(e) Gauss also used $(*)$ to show that $(2/p) = (-1)^{(p^2-1)/8}$. Adapt the argument given above to prove this. Hint: when $p \equiv 3,5 \bmod 8$, show that p is properly represented by a form of discriminant 8. When $p \equiv 1 \bmod 8$, note that the form $2x^2 + xy + ((1-p)/8)y^2$ has discriminant p and represents 2, and the argument is similar when $p \equiv 7 \bmod 8$.

3.14. Use Gauss' definition of genus to divide the forms of discriminant -164 into genera. Hint: the forms are given in (2.33). Notice that this is much easier than working with our original definition!

3.15. Prove (vi) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) and (vi) $\Rightarrow$ (v) $\Rightarrow$ (i) of Theorem 3.21.

3.16. Prove that the forms $x^2 + 18y^2$ and $2x^2 + 9y^2$ are rationally equivalent but belong to different genera. Hint: if they represent the same values in $(\mathbb{Z}/72\mathbb{Z})^*$, then the same is true for any divisor of 72.

3.17. Let D be a field discriminant, i.e., $D \equiv 1 \bmod 4$, D squarefree, or $D = 4k$, $k \not\equiv 1 \bmod 4$, k squarefree. Let $f(x,y)$ and $g(x,y)$ be rationally equivalent forms of discriminant D. We want to prove that they lie in the same genus.

(a) Let m be prime to D and represented by $g(x,y)$. Show that $f(x,y)$ represents d^2m for some nonzero integer d.

(b) Show that $f(x,y)$ and $g(x,y)$ lie in the same genus. Hint: by Exercise 2.1, $f(x,y)$ properly represents m' where $d'^2m' = d^2m$ for some integer d'. Show that m' is relatively prime to D. To do this, use Lemma 2.3 to write $f(x,y) = m'x^2 + bxy + cy^2$.

3.18. When $D = -4n$ is a field discriminant, we can use Theorem 3.21 to give a different proof that every form in the principal genus is a square (this is part (ii) of Theorem 3.15). Let $f(x,y)$ be a form of discriminant $-4n$ which lies in the principal genus.

(a) Show that $f(x,y)$ properly represents a number of the form a^2, where a is odd and relatively prime to n. Hint: use part (v) of Theorem 3.21.

(b) By (a), we may assume that $f(x,y) = a^2x^2 + 2bxy + cy^2$. Show that $\gcd(a,2b) = 1$, and conclude that $g(x,y) = ax^2 + 2bxy + acy^2$ has relatively prime coefficients and discriminant $-4n$.

(c) Show that $f(x,y)$ is the Dirichlet composition of $g(x,y)$ with itself.

This argument is due to Arndt (see Smith [95, pp. 254–256]), though Arndt proved (a) using the theorem of Legendre discussed in Exercise 2.24. Note that (a) can be restated in terms of ternary forms: if $f(x,y)$ is in the principal genus, then (a) proves that the ternary form $f(x,y) - z^2$ has a nontrivial zero. This result shows that there is a connection between ternary forms and genus theory. It is therefore not surprising that Gauss used ternary forms in his proof of Theorem 3.15.

3.19. Let $C(D)$ be the class group of forms of discriminant $D < 0$. Prove that the following statements are equivalent:

(i) Every genus of discriminant D consists of a single class.

(ii) $C(D) \simeq \{\pm 1\}^{\mu-1}$, where μ is as in Proposition 3.11.

(iii) Every genus of discriminant D consists of equivalent forms.

3.20. In this exercise we will prove Lemma 3.25. Let $m > 0$ be odd and prime to $n > 1$.

(a) Show that the number of solutions modulo m of the congruence

$$x^2 \equiv -n \bmod m$$

is given by the formula

$$\prod_{p|m} \left(1 + \left(\frac{-n}{p}\right)\right).$$

(b) Consider forms $g(x,y)$ of discriminant $-4n$ of the form

$$g(x,y) = mx^2 + 2bxy + cy^2, \quad 0 \le b < m.$$

Show that the map sending $g(x,y)$ to $[b] \in (\mathbb{Z}/m\mathbb{Z})^*$ induces a bijection between the $g(x,y)$'s and the solutions modulo m of $x^2 \equiv -n \bmod m$.

(c) Let $f(x,y)$ have discriminant $-4n$ and let $f(u,v) = m$ be a proper representation. Pick r_0, s_0, so that $us_0 - vr_0 = 1$, and set $r = r_0 + uk$, $s = s_0 + vk$. Note that as $k \in \mathbb{Z}$ varies, we get all solutions of $us - vr = 1$. Then set

$$g(x,y) = f(ux + ry, vx + sy)$$

and show that there is a unique $k \in \mathbb{Z}$ such that $g(x,y)$ satisfies the condition of (b). This form is denoted $g_{u,v}(x,y)$.

(d) Show that the map sending a proper representation $f(u,v) = m$ to the form $g_{u,v}(x,y)$ is onto.

(e) If $g_{u',v'}(x,y) = g_{u,v}(x,y)$, let

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} u' & v' \\ r' & s' \end{pmatrix}^{-1} \begin{pmatrix} u & v \\ r & s \end{pmatrix}.$$

Show that $f(\alpha x + \beta y, \gamma x + \delta y) = f(x,y)$ and, since $n > 1$, show that $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right) = \pm\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. Hint: assume that $f(x,y)$ is reduced, and use the arguments from the uniqueness part of the proof of Theorem 2.8.

(f) Conclude that $g_{u',v'}(x,y) = g_{u,v}(x,y)$ if and only if $(u',v') = \pm(u,v)$, so that the map of (d) is exactly two-to-one. Combining this with (a) and (b), we get a proof of Lemma 3.25.

3.21. This exercise will use Lemma 3.25 to study the equation $m^3 = a^2 + 2b^2$.

(a) If m is odd, use Lemma 3.25 to show that the equations $m = x^2 + 2y^2$ and $m^3 = x^2 + 2y^2$ have the same number of proper solutions.

(b) If $m = a^2 + 2b^2$ is a proper representation, then show that

$$m^3 = (a^3 - 6ab^2)^2 + 2(3a^2b - 2b^3)^2$$

is a proper representation.

(c) Show that the map sending (a,b) to $(a^3 - 6ab^2,\ 3a^2b - 2b^3)$ is injective. Hint: note that

$$(a + b\sqrt{-2})^3 = (a^3 - 6ab^2) + (3a^2b - 2b^3)\sqrt{-2}.$$

(d) Combine (a) and (c) to show that all proper representations of $m^3 = x^2 + 2y^2$, m odd, arise from (b).

3.22. Use Exercise 3.21 to prove Fermat's famous result that $(x,y) = (3, \pm 5)$ are the only integral solutions of the equation $x^3 = y^2 + 2$. Hint: first show that x must be odd, and then apply Exercise 3.21 to the proper representation $x^3 = y^2 + 2 \cdot 1^2$. It's likely that Fermat's original proof of this result was similar to the argument presented here, though he would have used a version of Lemma 1.4 to prove part (c) of Exercise 3.21. See Weil [106, pp. 68–69 and 71–73] for more details.

3.23. Let p be an odd prime of the form $x^2 + ny^2$, $n > 1$. Use Lemma 3.25 to show that the equation

$$p = x^2 + ny^2$$

has a unique solution once we require x and y to be nonnegative. Note also that Lemma 3.25 gives a very quick proof of Exercise 2.27.

3.24. This exercise will examine a generalization of the Jacobi symbol. Let P and Q be relatively prime nonzero integers, where Q is odd but possibly negative. Then define the extended Jacobi symbol (P/Q) via

$$\left(\frac{P}{Q}\right) = \begin{cases} (P/|Q|) & \text{when } |Q| > 1 \\ 1 & \text{when } |Q| = 1. \end{cases}$$

(a) Prove that when P and Q are odd and relatively prime, then

$$\left(\frac{P}{Q}\right)\left(\frac{Q}{P}\right) = (-1)^{(P-1)(Q-1)/4+(\mathrm{sgn}(P)-1)(\mathrm{sgn}(Q)-1)/4}$$

where $\mathrm{sgn}(P) = P/|P|$.

(b) Gauss' version of (a) is more complicated to state. First, given P and Q as above, he lets p denote the number of prime factors of Q (counted with multiplicity) for which P is not a quadratic residue. This relates to (P/Q) by the formula

$$\left(\frac{P}{Q}\right) = (-1)^p.$$

Interchanging P and Q, we get a similarly defined number q. To relate the parity of p and q, Gauss states a rule in [41, §133] which breaks up into 10 separate cases. Verify that the rule proved in (a) covers all 10 of Gauss' cases.

(c) Prove the supplementary laws:

$$\left(\frac{-1}{P}\right) = \mathrm{sgn}(P)(-1)^{(P-1)/2}$$
$$\left(\frac{2}{P}\right) = (-1)^{(P^2-1)/8}.$$

3.25. Let $p \equiv 1 \bmod 8$ be prime.

(a) If $C(-4p)$ is the class group of forms of discriminant $-4p$, then use genus theory to prove that

$$C(-4p) \simeq (\mathbb{Z}/2^a\mathbb{Z}) \times G$$

where $a \geq 1$ and G has odd order. Thus $2 \mid h(-4p)$.

(b) Let $f(x,y) = 2x^2 + 2xy + ((p+1)/2)y^2$. Use Gauss' definition of genus to show that $f(x,y)$ is in the principal genus.

(c) Use Theorem 3.15 to show that $C(-4p)$ has an element of order 4. Thus $4 \mid h(-4p)$.

§4. CUBIC AND BIQUADRATIC RECIPROCITY

In this section we will study cubic and biquadratic reciprocity and use them to prove Euler's conjectures for $p = x^2 + 27y^2$ and $p = x^2 + 64y^2$ (see (1.22) and (1.23)). An interesting feature of these reciprocity theorems is that each one requires that we extend the notion of integer: for cubic reciprocity we will use the ring

$$\mathbb{Z}[\omega] = \{a + b\omega : a, b \in \mathbb{Z}\}, \qquad \omega = e^{2\pi i/3} = (-1 + \sqrt{-3})/2, \tag{4.1}$$

and for biquadratic reciprocity we will use the Gaussian integers

$$\mathbb{Z}[i] = \{a + bi : a, b \in \mathbb{Z}\}, \qquad i = \sqrt{-1}. \tag{4.2}$$

Both $\mathbb{Z}[\omega]$ and $\mathbb{Z}[i]$ are subrings of the complex numbers (see Exercise 4.1). Our first task will be to describe the arithmetic properties of these rings and determine their units and primes. We will then define the generalized Legendre symbols $(\alpha/\pi)_3$ and $(\alpha/\pi)_4$ and state the laws of cubic and biquadratic reciprocity. The proofs will be omitted since excellent proofs are already available in print (see especially Ireland and Rosen [59, Chapter 9]). At the end of the section we will discuss Gauss' work on reciprocity and say a few words about the origins of class field theory.

A. $\mathbb{Z}[\omega]$ and Cubic Reciprocity

The law of cubic reciprocity is intimately bound up with the ring $\mathbb{Z}[\omega]$ of (4.1). The main tool used to study the arithmetic of $\mathbb{Z}[\omega]$ is the norm function: if $\alpha = a + b\omega$ is in $\mathbb{Z}[\omega]$, then its *norm* $N(\alpha)$ is the positive integer

$$N(\alpha) = \alpha\overline{\alpha} = a^2 - ab + b^2,$$

where $\overline{\alpha}$ is the complex conjugate of α (in Exercise 4.1 we will see that $\overline{\alpha} \in \mathbb{Z}[\omega]$). Note that the norm is multiplicative, i.e., for $\alpha, \beta \in \mathbb{Z}[\omega]$, we have

$$N(\alpha\beta) = N(\alpha)N(\beta)$$

(see Exercise 4.2). Using the norm, one can prove that $\mathbb{Z}[\omega]$ is a Euclidean ring:

Proposition 4.3. *Given $\alpha, \beta \in \mathbb{Z}[\omega]$, $\beta \neq 0$, there are γ, $\delta \in \mathbb{Z}[\omega]$ such that*

$$\alpha = \gamma\beta + \delta \qquad \textit{and} \qquad N(\delta) < N(\beta).$$

Thus $\mathbb{Z}[\omega]$ is a Euclidean ring.

Proof. The norm function $N(\alpha) = \alpha\overline{\alpha}$ is defined on $\mathbb{Q}(\omega) = \{r + s\omega : r, s \in \mathbb{Q}\}$ and satisfies $N(uv) = N(u)N(v)$ for $u, v \in \mathbb{Q}(\omega)$ (see Exercise 4.2). Then

$$\frac{\alpha}{\beta} = \frac{\alpha\overline{\beta}}{\beta\overline{\beta}} = \frac{\alpha\overline{\beta}}{N(\beta)} \in \mathbb{Q}(\omega),$$

so that $\alpha/\beta = r + s\omega$ for some $r, s \in \mathbb{Q}$. Let r_1, s_1 be integers such that $|r - r_1| \leq 1/2$ and $|s - s_1| \leq 1/2$, and then set $\gamma = r_1 + s_1\omega$ and $\delta = \alpha - \gamma\beta$. Note that $\gamma, \delta \in \mathbb{Z}[\omega]$ and $\alpha = \gamma\beta + \delta$. It remains to show that $N(\delta) < N(\beta)$. To see this, let $\epsilon = \alpha/\beta - \gamma = (r - r_1) + (s - s_1)\omega$, and note that

$$\delta = \alpha - \gamma\beta = \beta(\alpha/\beta - \gamma) = \beta\epsilon.$$

Since the norm is multiplicative, it suffices to prove that $N(\epsilon) < 1$. But

$$N(\epsilon) = N((r - r_1) + (s - s_1)\omega) = (r - r_1)^2 - (r - r_1)(s - s_1) + (s - s_1)^2,$$

and the desired inequality follows from $|r - r_1|$, $|s - s_1| \leq 1/2$. By the standard definition of a Euclidean ring (see, for example, Herstein [54, §3.7]), we are done. Q.E.D.

Corollary 4.4. *$\mathbb{Z}[\omega]$ is a PID (principal ideal domain) and a UFD (unique factorization domain).*

Proof. It is well known that any Euclidean ring is a PID and a UFD—see, for example, Herstein [54, Theorems 3.7.1 and 3.7.2]. Q.E.D.

For completeness, let's recall the definitions of PID and UFD. Let R be an integral domain. An ideal of R is *principal* if it can be written in the form $\alpha R = \{\alpha\beta : \beta \in R\}$ for some $\alpha \in R$, and R is a PID if every ideal of R is principal. To explain what a UFD is, we first need to define units, associates and irreducibles:

(i) $\alpha \in R$ is a *unit* if $\alpha\beta = 1$ for some $\beta \in R$.

(ii) $\alpha, \beta \in R$ are *associates* if α is a unit times β. This is equivalent to $\alpha R = \beta R$.

(iii) A nonunit $\alpha \in R$ is *irreducible* if $\alpha = \beta\gamma$ in R implies that β or γ is a unit.

Then R is a UFD if every nonunit $\alpha \neq 0$ can be written as a product of irreducibles, and given two such factorizations of α, each irreducible in the first factorization can be matched up in a one-to-one manner with an associate irreducible in the second. Thus factorization is unique up to order and associates.

It turns out that being a PID is the stronger property: every PID is a UFD (see Ireland and Rosen [59, §1.3]), but the converse is not true (see Exercise 4.3). Given a nonunit $\alpha \neq 0$ in a PID R, the following statements are equivalent:

(i) α is irreducible.

(ii) α is prime (an element α of R is *prime* if $\alpha \mid \beta\gamma$ implies $\alpha \mid \beta$ or $\alpha \mid \gamma$).

(iii) αR is a prime ideal (an ideal $\mathfrak{p}$ of R is *prime* if $\beta\gamma \in \mathfrak{p}$ implies $\beta \in \mathfrak{p}$ or $\gamma \in \mathfrak{p}$).

(iv) αR is a maximal ideal.

(See Exercise 4.4 for the proof.)

Since $\mathbb{Z}[\omega]$ is a PID and a UFD, the next step is to determine the units and primes of $\mathbb{Z}[\omega]$. Let's start with the units:

Lemma 4.5.

(i) *An element $\alpha \in \mathbb{Z}[\omega]$ is a unit if and only if $N(\alpha) = 1$.*

(ii) *The units of $\mathbb{Z}[\omega]$ are $\mathbb{Z}[\omega]^* = \{\pm 1, \pm\omega, \pm\omega^2\}$.*

Proof. See Exercise 4.5. Q.E.D.

The next step is to describe the primes of $\mathbb{Z}[\omega]$. The following lemma will be useful:

Lemma 4.6. *If $\alpha \in \mathbb{Z}[\omega]$ and $N(\alpha)$ is a prime in $\mathbb{Z}$, then α is prime in $\mathbb{Z}[\omega]$.*

Proof. Since $\mathbb{Z}[\omega]$ is a PID, it suffices to prove that α is irreducible. So suppose that $\alpha = \beta\gamma$ in $\mathbb{Z}[\omega]$. Taking norms, we obtain the integer equation

$$N(\alpha) = N(\beta\gamma) = N(\beta)N(\gamma)$$

(recall that the norm is multiplicative). Since $N(\alpha)$ is prime by assumption, this implies that $N(\beta)$ or $N(\gamma)$ is 1, so that β or γ is a unit by Lemma 4.5. Q.E.D.

We can now determine all primes in $\mathbb{Z}[\omega]$:

Proposition 4.7. *Let p be a prime in $\mathbb{Z}$. Then:*

(i) *If $p = 3$, then $1 - \omega$ is prime in $\mathbb{Z}[\omega]$ and $3 = -\omega^2(1-\omega)^2$.*

(ii) *If $p \equiv 1 \bmod 3$, then there is a prime $\pi \in \mathbb{Z}[\omega]$ such that $p = \pi\overline{\pi}$, and the primes π and $\overline{\pi}$ are nonassociate in $\mathbb{Z}[\omega]$.*

(iii) *If $p \equiv 2 \bmod 3$, then p remains prime in $\mathbb{Z}[\omega]$.*

Furthermore, every prime in $\mathbb{Z}[\omega]$ is associate to one of the primes listed in (i)–(iii) *above.*

Proof. Since $N(1-\omega) = 3$, Lemma 4.6 implies that $1-\omega$ is prime in $\mathbb{Z}[\omega]$, and (i) follows. To prove (ii), suppose that $p \equiv 1 \bmod 3$. Then $(-3/p) = 1$, so that p is represented by a reduced form of discriminant -3 (this is Theorem 2.16). The only such form is $x^2 + xy + y^2$, so that p can be written as $a^2 - ab + b^2$. Then $\pi = a + b\omega$ and $\overline{\pi} = a + b\omega^2$ have norms $N(\pi) = N(\overline{\pi}) = p$ and hence are prime in $\mathbb{Z}[\omega]$ by Lemma 4.6. In Exercise 4.7 we will prove that π and $\overline{\pi}$ are nonassociate. The proof of (iii) is left to the reader (see Exercise 4.7).

It remains to show that all primes in $\mathbb{Z}[\omega]$ are associate to one of the above. Let's temporarily call the primes given in (i)–(iii) the *known primes* of $\mathbb{Z}[\omega]$, and let α be any prime of $\mathbb{Z}[\omega]$. Then $N(\alpha) = \alpha\overline{\alpha}$ is an ordinary integer and may be factored into integer primes. But (i)–(iii) imply that any integer prime is a product of known primes in $\mathbb{Z}[\omega]$, and consequently $\alpha\overline{\alpha} = N(\alpha)$ is also a product of known primes. The proposition then follows since $\mathbb{Z}[\omega]$ is a UFD. Q.E.D.

Given a prime π of $\mathbb{Z}[\omega]$, we get the maximal ideal $\pi\mathbb{Z}[\omega]$ of $\mathbb{Z}[\omega]$. The quotient ring $\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega]$ is a thus a field. We can describe this field more carefully as follows:

Lemma 4.8. *If π is a prime of $\mathbb{Z}[\omega]$, then the quotient field $\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega]$ is a finite field with $N(\pi)$ elements. Furthermore, $N(\pi) = p$ or p^2 for some integer prime p, and:*

(i) *If $p = 3$ or $p \equiv 1 \bmod 3$, then $N(\pi) = p$ and $\mathbb{Z}/p\mathbb{Z} \simeq \mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega]$.*

(ii) *If $p \equiv 2 \bmod 3$, then $N(\pi) = p^2$ and $\mathbb{Z}/p\mathbb{Z}$ is the unique subfield of order p of the field $\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega]$ of p^2 elements.*

Proof. In §7 we will prove that if π is a nonzero element of $\mathbb{Z}[\omega]$, then $\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega]$ is a finite ring with $N(\pi)$ elements (see Lemma 7.14 or Ireland and Rosen [59, §§9.2 and 14.1]). Then (i) and (ii) follow easily (see Exercise 4.8). Q.E.D.

Given α, β and π in $\mathbb{Z}[\omega]$, we will write $\alpha \equiv \beta \bmod \pi$ to indicate that α and β differ by a multiple of π, i.e., that they give the same element in $\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega]$. Using this notation, Lemma 4.8 gives us the following analog of Fermat's Little Theorem:

Corollary 4.9. *If π is prime in $\mathbb{Z}[\omega]$ and doesn't divide $\alpha \in \mathbb{Z}[\omega]$, then*

$$\alpha^{N(\pi)-1} \equiv 1 \bmod \pi.$$

Proof. This follows because $(\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega])^*$ is a finite group with $N(\pi) - 1$ elements. Q.E.D.

Given these properties of $\mathbb{Z}[\omega]$, we can now define the generalized Legendre symbol $(\alpha/\pi)_3$. Let π be a prime of $\mathbb{Z}[\omega]$ not dividing 3 (i.e., not associate to $1-\omega$). It is straightforward to check that $3 \mid N(\pi) - 1$ (see Exercise 4.9). Now suppose that $\alpha \in \mathbb{Z}[\omega]$ is not divisible by π. It follows from Corollary 4.9 that $x = \alpha^{(N(\pi)-1)/3}$ is a root of $x^3 \equiv 1 \bmod \pi$. Since

$$x^3 - 1 \equiv (x-1)(x-\omega)(x-\omega^2) \bmod \pi$$

and π is prime, it follows that

$$\alpha^{(N(\pi)-1)/3} \equiv 1, \omega, \omega^2 \bmod \pi.$$

However, the cube roots of unity $1, \omega, \omega^2$ are incongruent modulo π. To see this, note that if any two were congruent, then we would have $1 \equiv \omega \bmod \pi$, which would contradict π not associate to $1-\omega$ (see Exercise 4.9 for the details). Then we define the *Legendre symbol* $(\alpha/\pi)_3$ to be the unique cube root of unity such that

$$\alpha^{(N(\pi)-1)/3} \equiv \left(\frac{\alpha}{\pi}\right)_3 \bmod \pi. \tag{4.10}$$

The basic properties of the Legendre symbol are easy to work out. First, from (4.10), one can show

$$\left(\frac{\alpha\beta}{\pi}\right)_3 = \left(\frac{\alpha}{\pi}\right)_3 \left(\frac{\beta}{\pi}\right)_3,$$

and second, $\alpha \equiv \beta \bmod \pi$ implies that

$$\left(\frac{\alpha}{\pi}\right)_3 = \left(\frac{\beta}{\pi}\right)_3$$

(see Exercise 4.10). The Legendre symbol may thus be regarded as a group homomorphism from $(\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega])^*$ to $\mathbb{C}^*$.

An important fact is that the multiplicative group of any finite field is cyclic (see Ireland and Rosen [59, §7.1]). In particular, $(\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega])^*$ is cyclic, which implies that

$$\begin{aligned} \left(\frac{\alpha}{\pi}\right)_3 = 1 &\iff \alpha^{(N(\pi)-1)/3} \equiv 1 \bmod \pi \\ &\iff x^3 \equiv \alpha \bmod \pi \quad \text{has a solution in } \mathbb{Z}[\omega] \end{aligned} \tag{4.11}$$

(see Exercise 4.11). This establishes the link between the Legendre symbol and cubic residues. Note that one-third of $(\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega])^*$ consists of cubic residues (where the Legendre symbol equals 1), and the remaining two-thirds consist of nonresidues (where the symbol equals ω or ω^2). Later on we will explain how this relates to the more elementary notion of cubic residues of integers.

To state the law of cubic reciprocity, we need one final definition: a prime π is called *primary* if $\pi \equiv \pm 1 \bmod 3$. Given any prime π not dividing 3, one can show that exactly two of the six associates $\pm\pi$, $\pm\omega\pi$ and $\pm\omega^2\pi$ are primary (see Exercise 4.12). Then the law of cubic reciprocity states the following:

Theorem 4.12. *If π and θ are primary primes in $\mathbb{Z}[\omega]$ of unequal norm, then*

$$\left(\frac{\theta}{\pi}\right)_3 = \left(\frac{\pi}{\theta}\right)_3.$$

Proof. See Ireland and Rosen [59, §§9.4–9.5] or Smith [95, pp. 89–91]. Q.E.D.

Notice how simple the statement of the theorem is—it's among the most elegant of all reciprocity theorems (biquadratic reciprocity, to be stated below, is a bit more complicated). The restriction to primary primes is a normalization analogous to the normalization $p > 0$ that we make for ordinary primes. Some books (such as Ireland and Rosen [59]) define primary to mean $\pi \equiv -1 \bmod 3$. Since $(-1/\pi)_3 = 1$, this doesn't affect the statement of cubic reciprocity.

There are also supplementary formulas for $(\omega/\pi)_3$ and $(1-\omega/\pi)_3$. Let π be prime and not associate to $1-\omega$. Then we may assume that $\pi \equiv -1 \bmod 3$ (if π is primary, one of $\pm\pi$ satisfies this condition). Writing $\pi = -1 + 3m + 3n\omega$, it can be

shown that

$$\left(\frac{\omega}{\pi}\right)_3 = \omega^{m+n}$$
$$\left(\frac{1-\omega}{\pi}\right)_3 = \omega^{2m}. \tag{4.13}$$

The first line of (4.13) is easy to prove (see Exercise 4.13), while the second is more difficult (see Ireland and Rosen [59, p. 114] or Exercise 9.13).

Let's next discuss cubic residues of integers. If p is a prime, the basic question is: when does $x^3 \equiv a \bmod p$ have an integer solution? If $p = 3$, then Fermat's Little Theorem tells us that $a^3 \equiv a \bmod 3$ for all a, so that we always have a solution. If $p \equiv 2 \bmod 3$, then the map $a \mapsto a^3$ induces an automorphism of $(\mathbb{Z}/p\mathbb{Z})^*$ since $3 \nmid p-1$ (see Exercise 4.14), and consequently $x^3 \equiv a \bmod p$ is again always solvable. If $p \equiv 1 \bmod 3$, things are more interesting. In this case, $p = \pi\overline{\pi}$ in $\mathbb{Z}[\omega]$, and there is a natural isomorphism $\mathbb{Z}/p\mathbb{Z} \simeq \mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega]$ by Lemma 4.8. Thus, for $p \nmid a$, (4.11) implies that

$$x^3 \equiv a \bmod p \text{ is solvable in } \mathbb{Z} \iff \left(\frac{a}{\pi}\right)_3 = 1. \tag{4.14}$$

Furthermore, $(\mathbb{Z}/p\mathbb{Z})^*$ breaks up into three pieces of equal size, one of cubic residues and two of nonresidues.

We can now use cubic reciprocity to prove Euler's conjecture for primes of the form $x^2 + 27y^2$:

Theorem 4.15. *Let p be a prime. Then $p = x^2 + 27y^2$ if and only if $p \equiv 1 \bmod 3$ and 2 is a cubic residue modulo p.*

Proof. First, suppose that $p = x^2 + 27y^2$. This clearly implies that $p \equiv 1 \bmod 3$, so that we need only show that 2 is a cubic residue modulo p. Let $\pi = x + 3\sqrt{-3}y$, so that $p = \pi\overline{\pi}$ in $\mathbb{Z}[\omega]$. It follows that π is prime, and then by (4.14), 2 is a cubic residue modulo p if and only if $(2/\pi)_3 = 1$. However, both 2 and $\pi = x + 3\sqrt{-3}y$ are primary primes, so that cubic reciprocity implies

$$\left(\frac{2}{\pi}\right)_3 = \left(\frac{\pi}{2}\right)_3. \tag{4.16}$$

It thus suffices to prove that $(\pi/2)_3 = 1$. However, from (4.10), we know that

$$\left(\frac{\pi}{2}\right)_3 \equiv \pi \bmod 2 \tag{4.17}$$

since $(N(2)-1)/3 = 1$. So we need only show that $\pi \equiv 1 \bmod 2$. Since $\sqrt{-3} = 1 + 2\omega$, $\pi = x + 3\sqrt{-3}y = x + 3y + 6y\omega$, so that $\pi \equiv x + 3y \equiv x + y \bmod 2$. But x and y must have opposite parity since $p = x^2 + 27y^2$, and we are done.

Conversely, suppose that $p \equiv 1 \bmod 3$ is prime and 2 is a cubic residue modulo p. We can write p as $p = \pi\overline{\pi}$, and we can assume that π is a primary prime in $\mathbb{Z}[\omega]$. This means that $\pi = a + 3b\omega$ for some integers a and b. Thus

$$4p = 4\pi\overline{\pi} = 4(a^2 - 3ab + 9b^2) = (2a - 3b)^2 + 27b^2.$$

Once we show b is even, it will follow immediately that p is of the form $x^2 + 27y^2$.

We now can use our assumption that 2 is a cubic residue modulo p. From (4.14) we know that $(2/\pi)_3 = 1$, and then cubic reciprocity (4.16) tells us that $(\pi/2)_3 = 1$. But by (4.17), this implies $\pi \equiv 1 \bmod 2$, which we can write as $a + 3b\omega \equiv 1 \bmod 2$. This easily implies that a is odd and b is even, and $p = x^2 + 27y^2$ follows. The theorem is proved. Q.E.D.

B. $\mathbb{Z}[i]$ and Biquadratic Reciprocity

Our treatment of biquadratic reciprocity will be brief since the basic ideas are similar to what we did for cubic residues (for a complete discussion, see Ireland and Rosen [59, §§9.7–9.9]). Here, the appropriate ring is the ring of Gaussian integers $\mathbb{Z}[i]$ as defined in (4.2). The norm function $N(a+bi) = a^2 + b^2$ makes $\mathbb{Z}[i]$ into a Euclidean ring, and hence $\mathbb{Z}[i]$ is also a PID and a UFD. The analogs of Lemma 4.5 and 4.6 hold for $\mathbb{Z}[i]$, and it is easy to check that its units are ± 1 and $\pm i$ (see Exercise 4.16). The primes of $\mathbb{Z}[i]$ are described as follows:

Proposition 4.18. *Let p be a prime in $\mathbb{Z}$. Then:*

(i) *If $p = 2$, then $1 + i$ is prime in $\mathbb{Z}[i]$ and $2 = i^3(1+i)^2$.*

(ii) *If $p \equiv 1 \bmod 4$, then there is a prime $\pi \in \mathbb{Z}[i]$ such that $p = \pi\overline{\pi}$ and the primes π and $\overline{\pi}$ are nonassociate in $\mathbb{Z}[i]$.*

(iii) *If $p \equiv 3$ mod 4, then p remains prime in $\mathbb{Z}[i]$.*

Furthermore, every prime in $\mathbb{Z}[i]$ is associate to one of the primes listed in (i)–(iii) *above.*

Proof. See Exercise 4.16. Q.E.D.

We also have the following version of Fermat's Little Theorem: if π is prime in $\mathbb{Z}[i]$ and doesn't divide $\alpha \in \mathbb{Z}[i]$, then

$$\alpha^{N(\pi)-1} \equiv 1 \bmod \pi \tag{4.19}$$

(see Exercise 4.16).

These basic facts about the Gaussian integers appear in many elementary texts (e.g., Herstein [54, §3.8]), but such books rarely mention that the whole reason Gauss introduced the Gaussian integers was so that he could state biquadratic reciprocity. We will have more to say about this later.

We can now define the Legendre symbol $(\alpha/\pi)_4$. Given a prime $\pi \in \mathbb{Z}[i]$ not associate to $1 + i$, it can be proved that $\pm 1, \pm i$ are distinct modulo π and that $4 \mid N(\pi) - 1$

(see Exercise 4.17). Then, for α not divisible by π, the *Legendre symbol* $(\alpha/\pi)_4$ is defined to be the unique fourth root of unity such that

$$\alpha^{(N(\pi)-1)/4} \equiv \left(\frac{\alpha}{\pi}\right)_4 \bmod \pi. \tag{4.20}$$

As in the cubic case, we see that

$$\left(\frac{\alpha}{\pi}\right)_4 = 1 \iff x^4 \equiv \alpha \bmod \pi \text{ is solvable in } \mathbb{Z}[i],$$

and furthermore, the Legendre symbol gives a character from $(\mathbb{Z}[i]/\pi\mathbb{Z}[i])^*$ to $\mathbb{C}^*$, so that $(\mathbb{Z}[i]/\pi\mathbb{Z}[i])^*$ is divided into four equal parts (see Exercise 4.18). When $p \equiv 1 \bmod 4$, we have $p = \pi\overline{\pi}$ with $(\mathbb{Z}[i]/\pi\mathbb{Z}[i])^* \simeq (\mathbb{Z}/p\mathbb{Z})^*$, and the partition can be described as follows: one part consists of biquadratic residues (where the symbol equals 1), another consists of quadratic residues which aren't biquadratic residues (where the symbol equals -1), and the final two parts consist of quadratic nonresidues (where the symbol equals $\pm i$)—see Exercise 4.19.

A prime π of $\mathbb{Z}[i]$ is *primary* if $\pi \equiv 1 \bmod 2+2i$. Any prime not associate to $1+i$ has a unique associate which is primary (see Exercise 4.21). With this normalization, the law of biquadratic reciprocity can be stated as follows:

Theorem 4.21. *If π and θ are distinct primary primes in $\mathbb{Z}[i]$, then*

$$\left(\frac{\theta}{\pi}\right)_4 = \left(\frac{\pi}{\theta}\right)_4 (-1)^{(N(\theta)-1)(N(\pi)-1)/16}.$$

Proof. See Ireland and Rosen [59, §9.9] or Smith [95, pp. 76–37]. Q.E.D.

There are also supplementary laws which state that

$$\begin{aligned} \left(\frac{i}{\pi}\right)_4 &= i^{-(a-1)/2} \\ \left(\frac{1+i}{\pi}\right)_4 &= i^{(a-b-1-b^2)/4} \end{aligned} \tag{4.22}$$

where $\pi = a+bi$ is a primary prime. As in the cubic case, the first line of (4.22) is easy to prove (see Exercise 4.22), while the second is more difficult (see Ireland and Rosen [59, Exercises 32–37, p. 136]).

We can now prove Euler's conjecture about $p = x^2 + 64y^2$:

Theorem 4.23.

(i) *If $\pi = a+bi$ is a primary prime in $\mathbb{Z}[i]$, then*

$$\left(\frac{2}{\pi}\right)_4 = i^{ab/2}.$$

(ii) *If p is prime, then $p = x^2 + 64y^2$ if and only if $p \equiv 1 \bmod 4$ and 2 is a biquadratic residue modulo p.*

Proof. First note that (i) implies (ii). To see this, let $p \equiv 1 \bmod 4$ be prime. We can write $p = a^2 + b^2 = \pi\overline{\pi}$, where $\pi = a + bi$ is primary. Note that a is odd and b is even. Since $\mathbb{Z}/p\mathbb{Z} \simeq \mathbb{Z}[i]/\pi\mathbb{Z}[i]$, (i) shows that 2 is a biquadratic residue modulo p if and only if b is divisible by 8, and (ii) follows easily.

One way to prove (i) is via the supplementary laws (4.22) since $2 = i^3(1+i)^2$ (see Exercise 4.23). However, in 1857, Dirichlet found a proof of (i) that uses only quadratic reciprocity [27, Vol. II, pp. 261–262]. A version of this proof is given in Exercise 4.24 (see also Ireland and Rosen [59, Exercises 26–28, p. 64]). Q.E.D.

C. Gauss and Higher Reciprocity

Most of the above theorems were discovered by Gauss in the period 1805–1814, though the bulk of what he knew was never published. Only in 1828 and 1832, long after the research was completed, did Gauss publish his two memoirs on biquadratic residues [42, Vol. II, pp. 65–148] (see also [41, pp. 511–586, German editions] for a German translation). The first memoir treats the elementary theory of biquadratic residues of integers, and it includes a proof of Euler's conjecture for $x^2 + 64y^2$. In the second memoir. Gauss begins with a careful discussion of the Gaussian integers, and he explains their relevance to biquadratic reciprocity as follows [42, Vol. II, §30, p. 102]:

> the theorems on biquadratic residues gleam with the greatest simplicity and genuine beauty only when the field of arithmetic is extended to **imaginary** numbers, so that without restriction, the numbers of the form $a + bi$ constitute the object [of study], where as usual i denotes $\sqrt{-1}$ and the indeterminates a, b denote integral real numbers between $-\infty$ and $+\infty$. We will call such numbers **integral complex numbers** (numeros integros complexos) ...

Gauss' treatment of $\mathbb{Z}[i]$ includes most of what we did above, and in particular the terms norm, associate and primary are due to Gauss.

Gauss' statement of biquadratic reciprocity differs slightly from Theorem 4.21. In terms of the Legendre symbol, his version goes as follows: given distinct primary primes π and θ of $\mathbb{Z}[i]$,

If either π or θ is congruent to 1 modulo 4, then $(\pi/\theta)_4 = (\theta/\pi)_4$.

If both π and θ are congruent to $3 + 2i$ modulo 4, then $(\pi/\theta)_4 = -(\theta/\pi)_4$.

In Exercise 4.25 we will see that this is equivalent to Theorem 4.21. As might be expected, Gauss doesn't use the Legendre symbol in his memoir. Rather, he defines the *biquadratic character* of α with respect to π to be the number $\lambda \in \{0,1,2,3\}$ satisfying $\alpha^{(N(\pi)-1)/4} \equiv i^\lambda \bmod \pi$ (so that $(\alpha/\pi)_4 = i^\lambda$), and he states biquadratic reciprocity using the biquadratic character. For Gauss, this theorem is "the Fundamental Theorem of biquadratic residues" [42, Vol. II, §67, p. 138], but instead of giving a proof, Gauss comments that

> In spite of the great simplicity of this theorem, the proof belongs to the most hidden mysteries of higher arithmetic, and at least as things now stand, [the

> proof] can be explained only by the most subtle investigations, which would greatly exceed the limits of the present memoir.

Later on, we will have more to say about Gauss' proof.

In the second memoir, Gauss also makes his only published reference to cubic reciprocity [42, Vol. II, §30, p. 102]:

> The theory of cubic residues must be based in a similar way on a consideration of numbers of the form $a + bh$, where h is an imaginary root of the equation $h^3 - 1 = 0$, say $h = (-1 + \sqrt{-3})/2$, and similarly the theory of residues of higher powers leads to the introduction of other imaginary quantities.

So Gauss was clearly aware of the properties of $\mathbb{Z}[\omega]$, even if he never made them public.

Turning to Gauss' unpublished material, we find that one of the earliest fragments on higher reciprocity, dated around 1805, is the following "Beautiful Observation Made By Induction" [42, Vol. VIII, pp. 5 and 11]:

> 2 is a cubic residue or nonresidue of a prime number p of the form $3n + 1$, according to whether p is representable by the form $xx + 27yy$ or $4xx + 2xy + 7yy$.

This shows that Euler's conjecture for $x^2 + 27y^2$ was one of Gauss' starting points. And notice that Gauss was aware that he was separating forms in the same genus—the very problem we discussed in §3.

Around the same time, Gauss also rediscovered Euler's conjecture for $x^2 + 64y^2$ [42, Vol. X.1, p. 37]. But how did he come to make these conjectures? There are two aspects of Gauss' work that bear on this question. The first has to do with quadratic forms. Let's follow the treatment in Gauss' first memoir on biquadratic residues [42, Vol. II, §§12–14, pp. 75–78]. Let $p \equiv 1 \bmod 4$ be prime. If 2 is to be a biquadratic residue modulo p, it follows by quadratic reciprocity that $p \equiv 1 \bmod 8$ (see Exercise 4.26). By Fermat's theorem for $x^2 + 2y^2$, p can be written as $p = a^2 + 2b^2$, and Gauss proves the lovely result that 2 is a biquadratic residue modulo p if and only if $a \equiv \pm 1 \bmod 8$ (see Exercise 4.27). This is nice, but Gauss isn't satisfied:

> Since the decomposition of the number p into a single and double square is bound up so prominently with the classification of the number 2, it would be worth the effort to understand whether the decomposition into two squares, to which the number p is equally liable, perhaps promises a similar success.

Gauss then computes some numerical examples, and they show that when p is written as $a^2 + b^2$, 2 is a biquadratic residue exactly when b is divisible by 8. This could be how Gauss was led to the conjecture in the first place, and the same thing could have happened in the cubic case, where primes $p = 1 \bmod 3$ can be written as $a^2 + 3b^2$.

The cubic case most likely came first, for it turns out that Gauss describes a relation between $x^2 + 27y^2$ and cubic residues in the last section of *Disquisitiones*. This is where Gauss discusses the cyclotomic equation $x^p - 1 = 0$ and proves his celebrated theorem on the constructibility of regular polygons. To see what this has to do with cubic residues, let's describe a little of what he does. Given an odd prime p, let $\zeta_p = e^{2\pi i/p}$ be a primitive pth root of unity, and let g be a primitive root modulo p, i.e., g is an integer such that $[g]$ generates the cyclic group $(\mathbb{Z}/p\mathbb{Z})^*$. Now suppose that $p - 1 = ef$, and let λ be an integer. Gauss then defines [41, §343] the *period*

(f,λ) to be the sum

$$(f,\lambda)=\sum_{j=0}^{f-1}\zeta_p^{\lambda g^{ej}}.$$

These periods are the key to Gauss' study of the cyclotomic field $\mathbb{Q}(\zeta_p)$. If we fix f, then the periods $(f,1),(f,g),(f,g^2),\dots,(f,g^{e-1})$ are the roots of an irreducible integer polynomial of degree e, so that these periods are primitive elements of the unique subfield $\mathbb{Q}\subset K\subset\mathbb{Q}(\zeta_p)$ of degree e over $\mathbb{Q}$. See Cox [A7, Section 9.2] for more on Gauss' theory of periods.

When $p\equiv 1 \bmod 3$, we can write $p-1=3f$, and then the three above periods are $(f,1)$, (f,g) and (f,g^2). Gauss studies this case in [41, §358], and by analyzing the products of the periods, he deduces the amazing result that

$$\begin{aligned}&\text{If } 4p=a^2+27b^2 \text{ and } a\equiv 1 \bmod 3, \text{ then } N=p+a-2, \text{ where}\\ &N \text{ is the number of solutions modulo } p \text{ of } x^3-y^3\equiv 1 \bmod p.\end{aligned}\tag{4.24}$$

To see how cubic residues enter into (4.24), note that $N=9M+6$, where M is the number of nonzero cubic residues which, when increased by one, remain a nonzero cubic residue (see Exercise 4.29). Gauss conjectured this result in October 1796 and proved it in July 1797 [42, Vol. X.1, entries 39 and 67, pp. 505–506 and 519]. So Gauss was aware of cubic residues and quadratic forms in 1796. Gauss' proof of (4.24) is sketched in Exercise 4.29.

Statement (4.24) is similar to the famous last entry in Gauss' mathematical diary. In this entry, Gauss gives the following analog of (4.24) for the decomposition $p=a^2+b^2$ of a prime $p\equiv 1 \bmod 4$:

$$\begin{aligned}&\text{If } p=a^2+b^2 \text{ and } a+bi \text{ is primary, then } N=p-2a-3, \text{ where}\\ &N \text{ is the number of solutions modulo } p \text{ of } x^2+y^2+x^2y^2\equiv 1 \bmod p\end{aligned}$$

(see [42, Vol. X.1, entry 146, pp. 571–572]). In general, the study of the solutions of equations modulo p leads to the zeta function of a variety over a finite field. For an introduction to this extremely rich topic, see Ireland and Rosen [59, Chapter 11]. In §14 we will see how Gauss' results relate to elliptic curves with complex multiplication.

Going back to the cubic case, there is a footnote in [41, §358] which gives another interesting property of the periods $(f,1)$, (f,g) and (f,g^2):

$$\begin{aligned}&((f,1)+\omega(f,g)+\omega^2(f,g^2))^3=p(a+b\sqrt{-27})/2,\\ &\text{where } 4p=a^2+27b^2.\end{aligned}\tag{4.25}$$

The right-hand side is an integer in the ring $\mathbb{Z}[\omega]$, and one can show that $\pi=(a+b\sqrt{-27})/2$ is a primary prime in $\mathbb{Z}[\omega]$ and that $p=\pi\overline{\pi}$. This is how Gauss first encountered $\mathbb{Z}[\omega]$ in connection with cubic residues. Notice also that if we set $\chi(a)=(a/\pi)_3$ and pick the primitive root g so that $\chi(g)=\omega$, then

$$(f,1)+\omega(f,g)+\omega^2(f,g^2)=\sum_{a=1}^{p-1}\chi(a)\zeta_p^a.\tag{4.26}$$

This is an example of what we now call a cubic Gauss sum. See Ireland and Rosen [59, §§8.2–8.3] for the basic properties of Gauss sums and a modern treatment of (4.24) and (4.25).

The above discussion shows that Gauss was aware of cubic residues and $\mathbb{Z}[\omega]$ when he made his "Beautiful Observation" of 1805, and it's not surprising that two years later he was able to prove a version of cubic reciprocity [42, Vol. VIII, pp. 9–13]. The biquadratic case was harder, taking him until sometime around 1813 or 1814 to find a complete proof. We know this from a letter Gauss wrote Dirichlet in 1828, where Gauss mentions that he has possessed a proof of the "Main Theorem" for around 14 years [42, Vol. II, p. 516]. Exact dates are hard to come by, for most of the fragments Gauss left are undated, and it's not easy to match them up with his diary entries. For a fuller discussion of Gauss' work on biquadratic reciprocity, see Bachmann [42, Vol. X.2.1, pp. 52–60] or Rieger [84].

Gauss' proofs of cubic and biquadratic reciprocity probably used Gauss sums similar to (4.26), and many modern proofs run along the same lines (see Ireland and Rosen [59, Chapter 9]). Gauss sums were first used in Gauss' sixth proof of quadratic reciprocity (see [42, Vol. II, pp. 55–59] or [41, pp. 501–505, German editions]). This is no accident, for as Gauss explained in 1818:

> From 1805 onwards I have investigated the theory of cubic and biquadratic residues ... Theorems were found by induction ... which had a wonderful analogy with the theorems for quadratic residues. On the other hand, for a long time all attempts at complete proofs have been futile. This was the motive for endeavoring to add yet more proofs to those already known for quadratic residues, in the hope that of the many different methods given, one or the other would contribute to the illumination of the related arguments [for cubic and biquadratic-residues]. This hope was in no way in vain, for at last tireless labor has led to favorable success. Soon the fruit of this vigilance will be permitted to come to public light ...

(see [42, Vol. II, p. 50] or [41, p. 497, German editions]). The irony is that Gauss never did publish his proofs, and it was left to Eisenstein and Jacobi to give us the first complete treatments of cubic and biquadratic reciprocity (see Collinson [22] or Smith [95, pp. 76–92] for more on the history of these reciprocity theorems).

We will conclude this section with some remarks about what happened after Gauss. Number theory was becoming a much larger area of mathematics, and the study of quadratic forms and reciprocity laws began to diverge. In the 1830s and 1840s, Dirichlet introduced L-series and began the analytic study of quadratic forms, and simultaneously, Eisenstein and Jacobi worked out cubic and biquadratic reciprocity. Jacobi studied reciprocity for 5th, 8th and 12th powers, and Eisenstein proved octic reciprocity. Kummer was also studying higher reciprocity, and he introduced his "ideal numbers" to make up for the lack of unique factorization in $\mathbb{Q}(e^{2\pi i/p})$. Both he and Eisenstein were able to prove generalized reciprocity laws using these "ideal numbers" (see Ireland and Rosen [59, Chapter 14] and Smith [95, pp. 93–126]). In 1871 Dedekind made the transition from "ideal numbers" to ideals in rings of algebraic integers, laying the foundation for modern algebraic number theory and class field theory. Lemmermeyer's book [A15] contains a wealth of in-

formation about reciprocity in the nineteenth century. See also Chapter 8 of the book [A3] by Berndt, Evans and Williams.

But reciprocity was not the only force leading to class field theory: there was also complex multiplication. Euler, Lagrange, Legendre and others studied transformations of the elliptic integrals

$$\int \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}},$$

and they discovered that certain values of k, called *singular moduli*, gave elliptic integrals that could be transformed into complex multiples of themselves. This phenomenon came to be called *complex multiplication.* In working with complex multiplication, Abel observed that singular moduli and the roots of the corresponding transformation equations have remarkable algebraic properties. In modern terms, they generate *Abelian extensions* of $\mathbb{Q}(\sqrt{-n})$, i.e., Galois extensions of $\mathbb{Q}(\sqrt{-n})$ whose Galois group is Abelian. These topics will be discussed in more detail in Chapter Three.

Kronecker extended and completed Abel's work on complex multiplication, and in so doing he made the amazing conjecture that every Abelian extension of $\mathbb{Q}(\sqrt{-n})$ lies in one of the fields described above. Kronecker had earlier conjectured that every Abelian extension of $\mathbb{Q}$ lies in one of the cyclotomic fields $\mathbb{Q}(e^{2\pi i/n})$ (this is the famous Kronecker–Weber Theorem, to be proved in §8). Abelian extensions may seem far removed from reciprocity theorems, but Kronecker also noticed relations between singular moduli and quadratic forms. For example, his results on complex multiplication by $\sqrt{-31}$ led to the following corollary which he was fond of quoting:

$$p = x^2 + 31y^2 \iff \begin{cases} (x^3 - 10x)^2 + 31(x^2-1)^2 \equiv 0 \bmod p \\ \text{has an integral solution} \end{cases}$$

(see [68, Vol. II, pp. 93 and 99–100, Vol. IV, pp. 123–129]). This is similar to what we just proved for $x^2 + 27y^2$ and $x^2 + 64y^2$ using cubic and biquadratic reciprocity. So something interesting is going on here.

We thus have two interrelated questions of interest:

(i) Is there a general reciprocity law that subsumes the known ones?

(ii) Is there a general method for describing all Abelian extensions of a number field?

The crowning achievement of class field theory is that it solves both of these problems simultaneously: an Abelian extension L of a number field K is classified in terms of data intrinsic to K, and the key ingredient linking L to this data is the Artin Reciprocity Theorem. Complete statements of the theorems of class field theory will be given in Chapter Two, and in Chapter Three we will explain how complex multiplication is related to the class field theory of imaginary quadratic fields.

For a fuller account of the history of class field theory, see the article by W. and F. Ellison [32, §§III–IV] in Dieudonné's *Abrégé d'Histoire des Mathématiques 1700–1900.* Weil has a nice discussion of reciprocity and cyclotomic fields in [105] and

[107], and Edwards describes Kummer's "ideal numbers" in [31, Chapter 4]. See also Part I of Vlăduţ's book [A24] on Kronecker's Jugentraum.

D. Exercises

4.1. Prove that $\mathbb{Z}[\omega]$ and $Z[i]$ are subrings of the complex numbers and are closed under complex conjugation.

4.2. Let $\mathbb{Q}(\omega) = \{r+s\omega : r,s \in \mathbb{Q}\}$, and define the norm of $r+s\omega$ to be $N(r+s\omega) = (r+s\omega)\overline{(r+s\omega)}$.

(a) Show that $N(r+s\omega) = r^2 - rs + s^2$.

(b) Show that $N(uv) = N(u)N(v)$ for $u, v \in \mathbb{Q}(\omega)$.

4.3. It is well-known that $R = \mathbb{C}[x,y]$ is a UFD (see Herstein [54, Corollary 2 to Theorem 3.11.1]). Prove that $I = \{f(x,y) \in R : f(0,0) = 0\}$ is an ideal of R which is not principal, so that R is not a PID. Hint: $x, y \in I$.

4.4. Given a nonunit $\alpha \neq 0$ in a PID R, prove that α is irreducible $\iff$ α is prime $\iff$ αR is a prime ideal $\iff$ αR is a maximal ideal.

4.5. Prove Lemma 4.5. Hint for (ii): use (i) and (2.4).

4.6. While $\mathbb{Z}[\omega]$ is a PID and a UFD, this exercise will show that the closely related ring $\mathbb{Z}[\sqrt{-3}]$ has neither property.

(a) Show that ± 1 are the only units of $\mathbb{Z}[\sqrt{-3}]$.

(b) Show that 2, $1+\sqrt{-3}$ and $1-\sqrt{-3}$ are nonassociate and irreducible in $\mathbb{Z}[\sqrt{-3}]$. Since $4 = 2 \cdot 2 = (1+\sqrt{-3})(1-\sqrt{-3})$, these elements are not prime and thus $\mathbb{Z}[\sqrt{-3}]$ is not a UFD.

(c) Show that the ideal in $\mathbb{Z}[\sqrt{-3}]$ generated by 2 and $1+\sqrt{-3}$ is not principal. Thus $\mathbb{Z}[\sqrt{-3}]$ is not a PID.

4.7. This exercise is concerned with the proof of Proposition 4.7. Let p be a prime number.

(a) When $p \equiv 1 \bmod 3$, we showed that $p = \pi\overline{\pi}$ where π and $\overline{\pi}$ are prime in $\mathbb{Z}[\omega]$. Prove that π and $\overline{\pi}$ are nonassociate in $\mathbb{Z}[\omega]$.

(b) When $p \equiv 2 \bmod 3$, prove that p is prime in $\mathbb{Z}[\omega]$. Hint: show that p is irreducible. Note that by Lemma 2.5, the equation $p = N(\alpha)$ has no solutions.

4.8. Complete the proof of Lemma 4.8.

4.9. Let π be a prime of $\mathbb{Z}[\omega]$ not associate to $1-\omega$.

(a) Show that $3 \mid N(\pi) - 1$.

(b) If two of $1, \omega, \omega^2$ are congruent modulo π, then show that $1 \equiv \omega \bmod \pi$, and explain why this contradicts our assumption on π. This proves that 1, ω and ω^2 are distinct modulo π.

4.10. Let π be prime in $\mathbb{Z}[\omega]$, and let $\alpha, \beta \in \mathbb{Z}[\omega]$ be not divisible by π. Verify the following properties of the Legendre symbol.

(a) $(\alpha\beta/\pi)_3 = (\alpha/\pi)_3(\beta/\pi)_3$.

(b) $(\alpha/\pi)_3 = (\beta/\pi)_3$ when $\alpha \equiv \beta \bmod \pi$.

4.11. Let π be prime in $\mathbb{Z}[\omega]$. Assuming that $(\mathbb{Z}[\omega]/\pi\mathbb{Z}[\omega])^*$ is cyclic, prove (4.11).

4.12. Let π be a prime of $\mathbb{Z}[\omega]$ which is not associate to $1-\omega$. Prove that exactly two of the six associates of π are primary.

4.13. Prove the top line of (4.13).

4.14. Use the hints in the text to prove that the congruence $x^3 \equiv a \bmod p$ is always solvable when p is a prime congruent to 2 modulo 3.

4.15. In this problem we will give an application of cubic reciprocity which is similar to Theorem 4.15. Let $p \equiv 1 \bmod 3$ be a prime.

(a) Use the proof of Theorem 4.15 to show that $4p$ can be written in the form $4p = a^2 + 27b^2$, where $a \equiv 1 \bmod 3$. Conclude that $\pi = (a + 3\sqrt{-3}b)/2$ is a primary prime of $\mathbb{Z}[\omega]$ and that $p = \pi\overline{\pi}$.

(b) Show that the supplementary laws (4.13) can be written

$$\left(\frac{\omega}{\pi}\right)_3 = \omega^{2(a+2)/3}$$
$$\left(\frac{1-\omega}{\pi}\right)_3 = \omega^{(a+2)/3+b}$$

where π is as in part (a).

(c) Use (b) to show that $(3/\pi)_3 = \omega^{2b}$.

(d) Use (c) and (4.14) to prove that for a prime p,

$$4p = x^2 + 243y^2 \iff \begin{cases} p \equiv 1 \bmod 3 \text{ and } 3 \text{ is a} \\ \text{cubic residue modulo } p \end{cases}$$

Euler conjectured the result of (d) (in a slightly different form) in his *Tractatus* [33, Vol. V, pp. XXII and 250].

4.16. In this exercise we will discuss the properties of the Gaussian integers $\mathbb{Z}[i]$.

(a) Use the norm function to prove that $\mathbb{Z}[i]$ is Euclidean.

(b) Prove the analogs of Lemmas 4.5 and 4.6 for $\mathbb{Z}[i]$.

(c) Prove Proposition 4.18.

(d) Formulate and prove the analog of Lemma 4.8 for $\mathbb{Z}[i]$.

(e) Prove (4.19).

4.17. If π is a prime of $\mathbb{Z}[i]$ not associate to $1+i$, show that $4 \mid N(\pi)-1$ and that ± 1 and $\pm i$ are all distinct modulo π.

4.18. This exercise is devoted to the properties of the Legendre symbol $(\alpha/\pi)_4$, where π is prime in $\mathbb{Z}[i]$ and α is not divisible by π.

(a) Show that $\alpha^{(N(\pi)-1)/4}$ is congruent to a unique fourth root of unity modulo π. This shows that the Legendre symbol, as given in (4.20), is well-defined. Hint: use Exercise 4.17.

(b) Prove that the analogs of the properties given in Exercise 4.10 hold for $(\alpha/\pi)_4$.

(c) Prove that

$$\left(\frac{\alpha}{\pi}\right)_4 = 1 \iff x^4 \equiv \alpha \bmod \pi \quad \text{is solvable in } \mathbb{Z}[i].$$

4.19. In this exercise we will study the integer congruence $x^4 \equiv a \bmod p$, where $p \equiv 1 \bmod 4$ is prime and a is an integer not divisible by p.

(a) Write $p = \pi\overline{\pi}$ in $\mathbb{Z}[i]$. Then use (4.20) to show that $(a/\pi)_4{}^2 = (a/p)$, and conclude that $(a/\pi)_4 = \pm 1$ if and only if $(a/p) = 1$.

(b) Verify the partition of $(\mathbb{Z}/p\mathbb{Z})^*$ described in the discussion following (4.20).

4.20. Here we will study the congruence $x^4 \equiv a \bmod p$ when $p \equiv 3 \bmod 4$ is prime and a is an integer not divisible by p.

(a) Use (4.20) to show that $(a/p)_4 = 1$. Thus a is a fourth power modulo p in the ring $\mathbb{Z}[i]$.

(b) Show that a is the biquadratic residue of an *integer* modulo p if and only if $(a/p) = 1$. Hint: study the maps $\phi_k(x) = x^{2^k}$ on an Abelian group of order $2m$, m odd.

4.21. If a prime π of $\mathbb{Z}[i]$ is not associate to $1+i$, then show that a unique associate of π is primary.

4.22. Prove the top formula of (4.22).

4.23. Use the supplementary laws (4.22) to prove part (i) of Theorem 4.23.

4.24. Let $p \equiv 1 \bmod 4$ be prime, and write $p = a^2 + b^2$, where a is odd and b is even. The goal of this exercise is to present Dirichlet's elementary proof that $(2/\pi)_4 = i^{ab/2}$, where $\pi = a + bi$.

(a) Use quadratic reciprocity for the Jacobi symbol to prove that $(a/p) = 1$.

(b) Use $2p = (a+b)^2 + (a-b)^2$ and quadratic reciprocity to show that

$$\left(\frac{a+b}{p}\right) = (-1)^{((a+b)^2-1)/8}.$$

(c) Use (b) and (4.20) to show that

$$\left(\frac{a+b}{p}\right) = \left(\frac{i}{\pi}\right)_4 i^{ab/2}.$$

Hint: $-1 = i^2$.

(d) From $(a+b)^2 \equiv 2ab \bmod p$, deduce that

(i) $(a+b)^{(p-1)/2} \equiv (2ab)^{(p-1)/4} \bmod p$.

(ii) $(a+b/p) = (2ab/\pi)_4$.

(e) Show that $2ab \equiv 2a^2 i \bmod \pi$, and then use (a) and Exercise 4.19 to show that

$$\left(\frac{2ab}{\pi}\right)_4 = \left(\frac{2i}{\pi}\right)_4.$$

(f) Combine (c), (d) and (e) to show that $(2/\pi)_4 = i^{ab/2}$.

4.25. In this exercise we will study Gauss' statement of biquadratic reciprocity.

(a) If π is a primary prime of $\mathbb{Z}[i]$, then show that either $\pi \equiv 1 \bmod 4$ or $\pi \equiv 3 + 2i \bmod 4$.

(b) Let π and θ be distinct primary primes in $\mathbb{Z}[i]$. Show that biquadratic reciprocity is equivalent to the following two statements:

If either π or θ is congruent to 1 modulo 4, then $(\pi/\theta)_4 = (\theta/\pi)_4$.

If π and θ are both congruent to $3+2i$ modulo 4, then $(\pi/\theta)_4 = -(\theta/\pi)_4$.

This is how Gauss states biquadratic reciprocity in [42, Vol. II, §67, p. 138].

4.26. If 2 is a biquadratic residue modulo an odd prime p, prove that $p \equiv \pm 1 \bmod 8$.

4.27. In this exercise, we will present Gauss' proof that for a prime $p \equiv 1 \bmod 8$, the biquadratic character of 2 is determined by the decomposition $p = a^2 + 2b^2$. As usual, we write $p = \pi\overline{\pi}$ in $\mathbb{Z}[i]$.

(a) Show that $(-1/\pi)_4 = 1$ when $p \equiv 1 \bmod 8$.

(b) Use the properties of the Jacobi symbol to show that

$$\left(\frac{a}{p}\right) = (-1)^{(a^2-1)/8}.$$

(c) Use the Jacobi symbol to show that $(b/p) = 1$. Hint: write $b = 2^m c$, c odd, and first show that $(c/p) = 1$.

(d) Show that

$$\left(\frac{2}{\pi}\right)_4 = \left(\frac{-2b^2}{\pi}\right)_4 = \left(\frac{a^2}{\pi}\right)_4 = \left(\frac{a}{p}\right).$$

Hint: use Exercise 4.19.

Combining (c) and (d), we see that $(2/\pi_4) = (-1)^{(a^2-1)/8}$, and Gauss' claim follows. If you read Gauss' original argument [42, Vol. II, §13], you'll appreciate how much the Jacobi symbol simplifies things.

4.28. Let (f,λ) and (f,μ) be periods, and write $(f,\mu) = \zeta^{\mu_1} + \cdots + \zeta^{\mu_f}$. Then prove that

$$(f,\lambda)\cdot(f,\mu) = \sum_{j=1}^{f}(f,\lambda+\mu_j).$$

4.29. Let $p \equiv 1 \bmod 3$ be prime, and set $p-1 = 3f$. Let $(f,1)$, (f,g) and (f,g^2) be the periods as in the text. Recall that g is a primitive root modulo p. In this problem we will describe Gauss' proof of (4.24) (see [41, §358]). For i, $j \in \{0,1,2\}$, let (ij) be the number of pairs (m,n), $0 \le m,n \le f-1$, such that

$$1 + g^{3m+i} \equiv g^{3n+j} \bmod p.$$

(a) Show that the number of solutions modulo p of the equation $x^3 - y^3 \equiv 1 \bmod p$ *is* $N = 9(00) + 6$.

(b) Use Exercise 4.28 to show that

$$(f,1)\cdot(f,1) = f + (00)(f,1) + (01)(f,g) + (02)(f,g^2)$$
$$(f,1)\cdot(f,g) = (10)(f,1) + (11)(f,g) + (12)(f,g^2)$$

and conclude that $(00)+(01)+(02) = f-1$ and $(10)+(11)+(12) = f$. Hint: $(f,0) = f$ and $-1 = (-1)^3$.

(c) Show that $(10) = (22), (11) = (20)$ and $(12) = (21)$. Hint: expand $(f,g)\cdot(f,1)$ and compare it to what you got in (b).

(d) Arguing as in (c), show that the 9 quantities (ij) reduce to three:

$$\alpha = (12) = (21) = (00) + 1$$
$$\beta = (01) = (10) = (22)$$
$$\gamma = (02) = (20) = (11).$$

(e) Note that $(f,1)\cdot(f,g)\cdot(f,g^2)$ is an integer. By expanding this quantity in terms of α, β and γ, show that

$$\alpha^2 + \beta^2 + \gamma^2 - \alpha = \alpha\beta + \beta\gamma + \alpha\gamma.$$

(f) Using (e), show that

$$(6\alpha - 3\beta - 3\gamma - 2)^2 + 27(\beta - \gamma)^2 = 12(\alpha + \beta + \gamma) + 4.$$

(g) Recall that $\alpha + \beta + \gamma = f$ (this was proved in (b)) and that $p - 1 = 3f$. Then use (f) to show that

$$4p = a^2 + 27b^2,$$

where $a = 6\alpha - 3\beta - 3\gamma - 2$ and $b = \beta - \gamma$.

(h) Let a be as in (g). Show that

$$a = 9\alpha - 3(\alpha + \beta + \gamma) - 2 = 9\alpha - p - 1.$$

Then use $\alpha = (00) + 1$ and (a) to conclude that

$$a = N - p + 2.$$

This proves (4.24).

In his first memoir on biquadratic residues [42, Vol. II, §§15–20, pp. 78–89], Gauss used a biquadratic analog of the (ij)'s (without any mention of periods) to determine the biquadratic character of 2.