LONDON MATHEMATICAL SOCIETY STUDENT TEXTS

Managing Editor: Professor C.M. SERIES, Mathematics Institute.
University of Warwick, Coventry CV4 7AL, United Kingdom

3 Local fields, J.W.S. CASSELS
4 An introduction to twistor theory: second edition, S.A. HUGGETT & K.P. TOD
5 Introduction to general relativity, L.P. HUGHSTON & K.P. TOD
7 The theory of evolution and dynamical systems, J. HOFBAUER & K. SIGMUND
8 Summing and nuclear norms in Banach space theory, G.J.O. JAMESON
9 Automorphisms of surfaces after Nielsen and Thurston, A. CASSON & S. BLEILER
11 Spacetime and singularities, G. NABER
12 Undergraduate algebraic geometry, MILES REID
13 An introduction to Hankel operators, J.R. PARTINGTON
15 Presentations of groups: second edition, D.L. JOHNSON
17 Aspects of quantum field theory in curved spacetime, S.A. FULLING
18 Braids and coverings: selected topics, VAGN LUNDSGAARD HANSEN
19 Steps in commutative algebra, R.Y. SHARP
20 Communication theory, C.M. GOLDIE & R.G.E. PINCH
21 Representations of finite groups of Lie type, FRANÇOIS DIGNE & JEAN MICHEL
22 Designs, graphs, codes, and their links, P.J. CAMERON & J.H. VAN LINT
23 Complex algebraic curves, FRANCES KIRWAN
24 Lectures on elliptic curves, J.W.S. CASSELS
25 Hyperbolic geometry, BIRGER IVERSEN
26 An introduction to the theory of L-functions and Eisenstein series, H. HIDA
27 Hilbert space: compact operators and the trace theorem, J.R. RETHERFORD
28 Potential theory in the complex plane, T. RANSFORD
29 Undergraduate commutative algebra, M. REID
31 The Laplacian on a Riemannian manifold, S. ROSENBERG
32 Lectures on Lie groups and Lie algebras, R. CARTER, G. SEGAL, & I. MACDONALD
33 A primer of algebraic D-modules, S.C. COUTINHO
34 Complex algebraic surfaces, A. BEAUVILLE
35 Young tableaux, W. FULTON
37 A mathematical introduction to wavelets, P. WOJTASZCZYK
38 Harmonic maps, loop groups, and integrable systems, M.A. GUEST
39 Set theory for the working mathematician, K. CIESIELSKI
40 Ergodic theory and dynamical systems, M. POLLICOTT & M. YURI
41 The algorithmic resolution of Diophantine equations, N.P. SMART
42 Equilibrium states in ergodic theory, G. KELLER
43 Fourier analysis on finite groups and applications, AUDREY TERRAS
44 Classical invariant theory, PETER J. OLVER

London Mathematical Society Student Texts 44

Classical Invariant Theory

Peter J. Olver
School of Mathematics
University of Minnesota

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011–4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1999

Typeface Computer Modern 10/12 pt. *System* TEX [AU]

*A catalog record of this book is available from
the British Library.*

Library of Congress Cataloging in Publication data
Olver, Peter J.
Classical invariant theory / Peter J. Olver
p. cm. – (London Mathematical Society student texts ; 44)
Includes bibliographical references and indexes.
ISBN 0 521 55243 5 (hc). – ISBN 0 521 55821 2 (pbk.)
1. Invariants. I. Title. II. Series.
QA201.O48 1999 98-33722
512.5 – dc21 CIP

0 521 55243 5 hardback
0 521 55821 2 paperpack

Transferred to digital printing 2003

To my parents — Grace E. Olver and Frank W.J. Olver

As all roads lead to Rome so I find in my own case at least that all algebraic inquiries, sooner or later, end at the Capitol of Modern Algebra over whose shining portal is inscribed the Theory of Invariants.

— Sylvester, quoted in [**72**; p. 143].

The *theory of invariants* came into existence about the middle of the nineteenth century somewhat like Minerva: a grown-up virgin, mailed in the shining armor of algebra, she sprang forth from Cayley's Jovian head. Her Athens over which she ruled and which she served as a tutelary and beneficent goddess was *projective geometry*.

— Weyl, [**230**].

Like the Arabian phoenix arising out of its ashes, the theory of invariants, pronounced dead at the turn of the century, is once again at the forefront of mathematics.

— Kung and Rota, [**135**].

Contents

Introduction

Classical invariant theory is the study of the intrinsic or geometrical properties of polynomials. This fascinating and fertile field was brought to life at the beginning of the last century just as the theory of solubility of polynomials was reaching its historical climax. It attained its zenith during the heyday of nineteenth-century mathematics, uniting researchers from many countries in a common purpose, and filling the pages of the foremost mathematical journals of the time. The dramatic and unexpected solution to its most fundamental problem — the finitude of the number of fundamental invariants — propelled the young David Hilbert into the position of the most renowned mathematician of his time. Following a subsequent decline, as more fashionable subjects appeared on the scene, invariant theory sank into obscurity during the middle part of this century, as the abstract approach entirely displaced the computational in pure mathematics. Ironically, though, its indirect influence continued to be felt in group theory and representation theory, while in abstract algebra the three most famous of Hilbert's general theorems — the Basis Theorem, the Syzygy Theorem, and the Nullstellensatz — were all born as lemmas (*Hilfsätze*) for proving "more important" results in invariant theory! Recent years have witnessed a dramatic resurgence of this venerable subject, with dramatic new applications, ranging from topology and geometry, to physics, continuum mechanics, and computer vision. This has served to motivate the dusting off of the old computational texts, while the rise of computer algebra systems has brought previously infeasible computations within our grasp. In short, classical invariant theory is the closest we come in mathematics to sweeping historical drama and romance. As a result, the subject should hold a particular fascination, not only for the student and practitioner, but also for any mathematician with a desire to understand the culture, sociology, and history of mathematics.

I wrote this introductory textbook in the hope of furthering the recent revival of classical invariant theory in both pure and applied mathematics. The presentation is not from an abstract, algebraic standpoint,

but rather as a subject of interest for applications in both mathematics and other scientific fields. My own training is in differential equations and mathematical physics, and so I am unashamed to restrict my attention to just real and complex polynomials. This approach allows me to directly employ differentiation and other analytical tools as the need arises. In this manner, the exposition at times resembles that of the classical texts from the last century, rather than that of more modern treatments that either presuppose an extensive training in the methods of abstract algebra or reduce the subject to a particular case of general tensor analysis. Nevertheless, a fair amount of more recent material and modern developments is covered, including several original results that have not appeared in print before. I have designed the text so that it can be profitably read by students having a fairly minimal number of mathematical prerequisites.

Notes to the Reader

The purpose of this book is to provide the student with a firm grounding in the basics of classical invariant theory. The text is written in a non-abstract manner and makes fairly low demands on the prospective reader. In addition, a number of innovations — in methodology, style, and actual results — have been included that should attract the attention of even the most well-seasoned researcher. We shall concentrate on the basic theory of binary forms, meaning polynomials in a single variable, under the action of the projective group of linear fractional transformations, although many of the methods and theoretical foundations to be discussed have far wider applicability. The classical constructions are all founded on the theory of groups and their representations, which are developed in detail from the beginning during the exposition.

The text begins with the easiest topic of all: the theory of a single real or complex quadratic polynomial in a single variable. Although completely elementary, this example encapsulates the entire subject and is well worth reviewing one more time — although an impatient reader can entirely omit this preliminary chapter. As any high school student knows, the solution to the quadratic equation relies on the associated discriminant. Less obvious is the fact that the discriminant is (relatively) unchanged under linear fractional transformations. Hence it forms the

first (both historical and mathematical) example of an *invariant* and so can be used for classification of canonical forms. The text starts in earnest in Chapter 2, which provides an overview of the basics of classical invariant theory within the context of binary forms. Here we meet up with the basic definitions of invariants and covariants, and investigate how the geometry of projective space governs the correspondence between homogeneous and inhomogeneous polynomials, as well as their transformation properties under, respectively, linear and projective transformations. The motivating examples of cubic and quartic polynomials are discussed in detail, including complete lists of invariants, covariants, and canonical forms. The Fundamental Theorem of Algebra guarantees the existence of a complete system of (complex) roots, whose geometrical configuration is governed by the invariants. Two particularly important invariants are the classical resultant, which indicates the existence of common roots to a pair of polynomials, and the discriminant, which indicates multiple roots of a single polynomial. The chapter concludes with a brief introduction to the Hilbert Basis Theorem, which states that every system of polynomials has only a finite number of polynomially independent invariants, along with remarks on the classification of algebraic relations or syzygies among the invariants.

With this preliminary survey as our motivating guide, the next two chapters provide a grounding in the modern mathematical foundations of the subject, namely, transformation groups and representation theory. Chapter 3 is a self-contained introduction to groups and their actions on spaces. Groups originally arose as the symmetries of a geometric or algebraic object; in our case the object is typically a polynomial. The chapter includes a discussion of the equivalence problem — when can two objects be transformed into each other by a suitable group element — and the allied concept of a canonical form. Chapter 4 concentrates on the theory of linear group actions, known as representations. For general transformation groups, the associated multiplier representations act on the functions defined on the space; the linear/projective actions on polynomials form a very particular instance of this general construction. The invariant functions arise as fixed points for such representations, and so the focus of classical invariant theory naturally falls within this general framework.

The next three chapters describe the core of the classical constructive algebraic theory of binary forms. The most important operations for producing covariants are the "transvection" processes, realized as

certain bilinear differential operators acting on binary forms, or, more generally, analytic functions. According to the First Fundamental Theorem of classical invariant theory, all of the invariants and covariants for any system of polynomials or, more generally, functions, can be constructed through iterated transvectants and, in the inhomogeneous case, scaling processes. Thus, a proper grounding in these basic techniques is essential. Traditionally, such invariant processes are based on the symbolic method, which is the most powerful computational tool for computing and classifying invariants. However, no aspect of the classical theory has been as difficult to formalize or as contentious. The point of view taken here is nonstandard, relying on the construction of covariants and invariants as differential polynomials. Taking inspiration from work of Gel'fand and Dikii, [**77**], in solitons and the formal calculus of variations, I introduce a transform that mimics the Fourier transform of classical analysis and maps differential polynomials into algebraic polynomials. The transform is, in essence, the symbolic method realized in a completely natural manner, applicable equally well to polynomials and more general functions. The chapter concludes with proofs of the First Fundamental Theorem, which states that every covariant has symbolic form given by a polynomial in certain "bracket factors", and the Second Fundamental Theorem, which completely classifies the syzygies among the brackets. Although the determination of a complete Hilbert basis for the covariants of a general binary form turns out to be an extremely difficult problem, which has been solved only for forms of low degree, I shall prove a result due to Stroh and Hilbert that constructs an explicit rational basis for a form of arbitrary degree.

Chapter 7 introduces a graphical version of the symbolic method that can be used to simply and pictorially analyze complicated invariant-theoretic identities for binary forms. Each symbolic expression has an equivalent directed graph, or "digraph" counterpart, whereby algebraic identities among the symbolic forms translate into certain graphical operations that bear much similarity to basic operations in knot theory, [**124**], and thereby lead to a significant simplification with visual appeal. As an application, I show how to implement Gordan's method for constructing a complete system of fundamental invariants and covariants for binary forms, illustrated by the cubic and quartic examples.

At this point, we have covered the classical algebraic techniques underlying the theory of binary forms. Since the group of linear/projective transformations depends analytically on parameters, it is an example of

a Lie transformation group. The theory of Lie groups includes a wide range of powerful calculus-based tools for the analysis of their invariants. Chapter 8 begins with a very brief introduction to Lie groups, including the general Frobenius Theorem that completely determines the local structure of the orbits and the fundamental invariants for regular actions. Here, invariants are classified up to functional dependence, rather than polynomial or rational dependence as was done in the more algebraic aspects of the theory; the number of fundamental invariants depends solely on the dimension of the group orbits. Even better, there is an explicit computation algorithm, which relies just on the Implicit Function Theorem, for constructing the invariants of regular Lie group actions. This method, known as "normalization", has its origins in Élie Cartan's theory of moving frames, [**33**, **93**], which was developed for studying the geometry of curves and surfaces. Surprisingly, the normalization method has not been developed at all in the standard literature; the construction relies on a new theory of moving frames for general transformation group actions recently established by the author in collaboration with Mark Fels, [**69**, **70**]. Applications to the classification of joint invariants and differential invariants for interesting transformation groups are provided.

In the theory of moving frames, the determination of symmetries, the complete solution to the equivalence problem, and the construction of canonical forms rely on the analysis of suitable differential invariants. In the case of planar curves, there is a single basic differential invariant — the group-theoretic *curvature* — along with a group-invariant *arc length* element. Higher order differential invariants are obtained by repeatedly differentiating curvature with respect to arc length. The first two fundamental differential invariants trace out the *signature set* which uniquely characterizes the curve up to group transformations. A direct application of the moving frame method leads to a remarkable theorem that the equivalence and symmetry of a binary form relies on merely *two* classical rational covariants! This result, first established in [**167**], reduces the entire complicated algebraic Hilbert basis to a simple pair of rational covariants whose functional dependencies completely encode the geometric properties of the binary form. I present a number of striking new consequences of this result, including a new bound on the number of discrete symmetries of polynomials. These innovative techniques are of much wider applicability and clearly deserve further development in the multivariate context.

While Chapter 8 develops "finite" Lie theory, the following chapter is concerned with Lie's powerful infinitesimal approach to invariance. Each Lie-theoretic object has an infinitesimal counterpart, and the replacement of complicated group-theoretic conditions by their infinitesimal analogs typically linearizes and significantly simplifies the analysis. The infinitesimal version of a Lie group is known as a Lie algebra, which contains the infinitesimal generators of the group action, realized as first order differential operators (or vector fields). Assuming connectivity, a function is invariant under the group if and only if it is annihilated by the infinitesimal generators, allowing methods from the theory of partial differential equations to be applied to the analysis of invariants. In the context of binary forms, the infinitesimal generators were, in fact, first recognized by Cayley, [**41**], to play an important role in the theory. I show how one can use these to build up general invariants from simpler "semi-" and "isobaric" invariants through an inductive procedure based on invariance under subgroups. The chapter culminates in a proof of the Hilbert Basis Theorem that relies on a particular differential operator that converts functions into invariants.

The final chapter is included to provide the reader with an orientation to pursue various generalizations of the basic methods and theories to multivariate polynomials and functions. Unfortunately, space has finally caught up with us at this point, and so the treatment is more superficial. Nevertheless, I hope that the reader will be sufficiently motivated to pursue the subject in more depth.

I have tried to keep the prerequisites to a minimum, so that the text can be profitably read by anyone trained in just the most standard undergraduate material. Certainly one should be familiar with basic linear algebra: vectors, matrices, linear transformations, Jordan canonical form, norms, and inner products — all of which can be found in any comprehensive undergraduate linear algebra textbook. Occasionally, I employ the tensor product construction. No knowledge of the general theory of polynomial equations is assumed. An introductory course in group theory could prove helpful to the novice but is by no means essential since I develop the theory of groups and their representations from scratch. All constructions take place over the real or complex numbers, and so no knowledge of more general field theory is ever required. One certainly does not need to take an abstract algebra course before starting; indeed, this text may serve as a good motivation or supplement for such a course!

In Chapters 8 and 9, I rely on multivariable differential calculus, at least as far as the Implicit Function Theorem, and the basic theory of first order systems of ordinary differential equations. In particular, the reader should be familiar with the solution to linear systems of differential equations, including matrix exponentials and their computation via Jordan canonical forms. I do not require any experience with Lie group theory or differential geometry, although the reader may wish to consult a basic text on manifolds, vector fields, and Lie groups to supplement the rather brief exposition here. (Chapter 1 of my own book [**168**] is particularly recommended!) Some of the more difficult results are stated without proof, although ample references are provided. I should remark that although the transform method adopted in Chapter 6 is inspired by the Fourier transform, no actual knowledge of the analytical Fourier theory is required.

Inevitably, the writing of an introductory text of moderate size requires making tough choices on what to include and what to leave out. Some of my choices are unorthodox. (Of course, if all choices were "orthodox", then there wouldn't be much point writing the book, as it would be a mere reworking of what has come before.) The most orthodox choice, followed in all the classical works as well as most modern introductions, is to concentrate almost entirely on the relatively modest realm of binary forms, relegating the vast hinterlands of multivariate polynomials and functions to an all too brief final chapter that cannot possibly do them justice. Of course, one motivation for this tactic is that most of the interesting explicit results and methods already make their appearance in the binary form case. Still, one tends to leave with the wish that such authors (including the present one) had more to say of substance in the multidimensional context.

Less orthodox choices include the reliance on calculus — differential operators, differential equations, differential invariants — as a framework for the general theory. Here we are in good company with the classics — Clebsch, [**49**], Gordan, [**89**], Grace and Young, [**92**], and even Hilbert, [**107**]. Post–Noetherian algebraists will no doubt become alarmed that I have regressed, in that the calculus-based tools are only valid in characteristic zero, or, more specifically, for the real and complex numbers, while "true" invariant theory requires that all fields be treated as equals, which means throwing out such "antiquated" analytical tools. My reply (and I speak here as the semi-applied mathematician I am) is that the primary physical and geometrical applications of invariant

theory, which, after all, motivated its development, remain either real or complex, and it is here that much of the depth, beauty, and utility of the subject still resides. Another, more provocative, response is that the more interesting generalization of the classical techniques is not necessarily to fields of nonzero characteristic, but rather to more general associative and non-associative algebras, starting with the quaternions, octonions, Clifford algebras, quantum groups, and so on. One retains calculus (the quaternion calculus is a particularly pretty case) but gives up commutativity (and even possibly associativity). The development of a non-commutative classical invariant theory remains, as far as I know, completely unexplored.

The most original inclusion is the application of the Cartan theory of moving frames to the determination of symmetries and a solution to the equivalence problem for binary forms. Most of the constructions and results in this part of the text are new but can be readily comprehended by an advanced undergraduate student. This connection between geometry and algebra, I believe, opens up new and extremely promising vistas in both subjects — not to mention the connections with computer vision and image processing that served as one of my original motivations.

An unorthodox omission is the combinatorial and enumerative techniques that receive a large amount of attention in most standard texts. This was a difficult decision, and a topic I really did want to include. However, as the length of the manuscript crept up and up, it became clear that something had to go, and I decided this was it. The combinatorial formulae that count the number of invariants, particularly those based on Hilbert and Molien series and their generating functions, are very pretty and well worth knowing; see [**200**, **204**], for instance. However, as far as practical considerations go, they merely serve as indicators of what to expect and are of less help in the actual determination and classification of invariants. Indeed, in all the examples presented here, enumeration formulae are never used, and so their omission will not leave any gaps in the exposition. But the reader is well advised to consult other sources to rectify this omission.

The text is designed for the active reader. As always, one cannot learn mathematics by merely reading or attending lectures — one needs to *do* mathematics in order to absorb it. Thus, a large number and variety of exercises, of varying degrees of difficulty, are liberally interspersed throughout the text. They either illustrate the general theory with additional interesting examples or supply further theoretical results of im-

portance that are left for the reader to verify. The student is strongly encouraged to attempt most exercises while studying the material.

I have also included many references and remarks of historical and cultural interest. I am convinced that one cannot learn a mathematical subject without being at least partially conversant with its roots and its original texts. Modern reformulations of classical mathematics, while sometimes (but not always) more digestible to the contemporary palate, often shortchange the contributions of the original masters. Worse yet, such rewritings can actually be harder for the novice to digest, since they tend to omit the underlying motivations or significance of the results and their interconnections with other parts of mathematics and applications. I am a firm believer in the need for a definite historical consciousness in mathematics. There is no better way of learning a theorem or construction than by going back to the original source, and a text (even at an introductory level) should make significant efforts to uncover and list where the significant ideas were conceived and brought to maturation. On the other hand, I do not pretend that my list of references is in any sense complete (indeed, the sheer volume of the nineteenth-century literature precludes almost any attempt at completeness); nevertheless, it includes many obscure but vital papers that clearly deserve a wider audience. I hope the reader is inspired to continue these historical and developmental studies in more depth.

The text has been typeset using the author's own OTEX system of macros. Details and software can be found at my web site:

`http://www.math.umn.edu/~olver` .

The figures were drawn with the aid of MATHEMATICA. Comments, corrections, and questions directed to the author are most welcome.

A Brief History

Classical invariant theory's origins are to be found in the early-nineteenth-century investigations by Boole, [**24**], into polynomial equations. The subject was nurtured by that indefatigable computer Cayley, to whom we owe many of the fundamental algorithms. Any reader of Cayley's collected works, [**36**], which include page after page of extensive explicit tables, cannot but be in awe of his computational stamina. (I often wonder what he might have accomplished with a functioning

computer algebra system!) While the British school, led by Cayley and the flamboyant Sylvester, joined by Hermite in France, was the first to plow the virgin land, the actual flowering and maturation of the theory passed over to the Germans. The first wave of German experts includes Aronhold, the progenitor of the mystical "symbolic method", Clebsch, whose contributions metamorphosed into basic formulae in representation theory with profound consequences for quantum physics, and, most prominently, Gordan, the first among equals. Gordan's crowning achievement was his computational procedure and proof of the fundamental Basis Theorem that guarantees only a finite number of independent invariants for any univariate polynomial. The classical references by Clebsch, [**49**], Faà di Bruno, [**67**], and Gordan, [**89**], describe the resulting invariant theory of binary forms. A very extensive history of the nineteenth-century invariant theory, including copious references, was written by F. Meyer, [**151**]. Modern historical studies by Fisher, [**72**], and Crilly, [**55**], also document the underlying sociological and cultural implications of its remarkable history.

Despite much effort, extending Gordan's result to polynomials in two or more variables proved too difficult, until, in a profound stroke of genius, David Hilbert dramatically unveiled his general Basis Theorem in 1890. Hilbert's first, existential proof has, of course, had an incomparable impact, not just in classical invariant theory, but in all of mathematics, since it opened the door to the abstract algebraic approach that has characterized a large fraction of twentieth-century mathematics. Its immediate impact was the discreditation of the once dominant computational approach, which gradually fell into disrepute. Only in recent years, with the advent of powerful computer algebra systems and a host of new applications, has the computational approach to invariant theory witnessed a revival.

Nevertheless, the dawn of the twentieth century saw the subject in full florescence, as described in the marvelous (and recently translated) lectures of Hilbert, [**107**]. The texts by Grace and Young, [**92**], and Elliott, [**65**], present the state of the computational art, while Weitzenböck, [**229**], reformulates the subject under the guiding light of the new physics and tensor analysis. So the popular version of history, while appealing in its drama, is not entirely correct; Hilbert's paper did not immediately kill the subject, but rather acted as a progressive illness, beginning with an initial shock, and slowly consuming the computational body of the theory from within, so that by the early 1920's the subject was clearly

moribund. Abstraction ruled: the disciples of Emmy Noether, a student of Gordan, led the fight against the discredited computational empire, perhaps as a reaction to Noether's original, onerous thesis topic that involved computing the invariants for a quartic form in three variables.

Although the classical heritage had vanished from the scene by mid-century, all was not quiet. The profoundly influential, yet often frustratingly difficult, book by Weyl, [**231**], places the classical theory within a much more general framework; polynomials now become particular types of tensorial objects, while, motivated by simultaneous developments in algebra and physics, the action of linear or linear fractional transformations is now extended to the vast realm of group representations. Attempts to reconcile both the classical heritage and Weyl's viewpoint with modern algebra and geometry have served to inspire a new generation of invariant theorists. Among the most influential has been Mumford's far-reaching development of the incisive methods of Hilbert, leading to the deep but abstract geometrical invariant theory, [**156**]. New directions, inspired by recent developments in representation theory and physics, appear in the recent work of Howe, [**112**]. Particular mention must be made of Rota and his disciples, [**94, 135**], whose efforts to place the less than rigorous classical theory, particularly the symbolic method, on a firm theoretical foundation have had significant influence. The comprehensive text of Gurevich, [**97**], is a particularly useful source, which helped inspire a vigorous, new Russian school of invariant theorists, led by Popov, [**181**], and Vinberg, [**226**], who have pushed the theory into fertile new areas.

Of course, one cannot fail to mention the rise of modern computer algebra. Even the masters of the last century became stymied by the sheer complexity of the algebraic formulas and manipulations that the subject breeds. The theory of Gröbner bases, cf. [**54**], has breathed new life into the computational aspect of the subject. Sturmfels' elegant book, [**204**], gives an excellent survey of current work in this direction and is particularly recommended to the student wishing to continue beyond the material covered here. The influence of classical invariant theory can be felt throughout mathematics and extends to significant physical applications, ranging from algebra and number theory, [**79**], through combinatorics, [**201**], Riemannian geometry, [**149, 150, 229**], algebraic topology, [**196**], and ordinary differential equations, [**235, 195**]. Applications include continuum mechanics, [**197**], dynamical systems, [**195**], engineering systems and control theory, [**213**], atomic physics, [**189**],

and even computer vision and image processing, [**157**]. This text should prove to be useful to students in all of these areas and many more.

Acknowledgments

Many people deserve thanks for helping inspire my interest in this subject and desire to put pen to paper (or, more accurately, finger to keyboard). They include John Ball, whose questions in nonlinear elasticity directly motivated my initial forays; Gian–Carlo Rota, whose wonderful lectures opened my vistas; and Bernd Sturmfels, who patiently introduced me to modern computational tools. I would particularly like to thank my student Irina Berchenko for her careful proofreading of the manuscript. Most of all, I must express my heartfelt gratitude to my wonderful wife, Cheri Shakiban, who directly collaborated with me on several invariant theoretic papers, [**171**, **172**], which form the basis of significant parts of this text. I am incredibly fortunate that she has been such a major part of my life.

I have dedicated this book to my parents. My father, Frank W.J. Olver, is a great applied analyst and certainly played a direct, inspirational role in my choice of career. I wish my mother, Grace E. Olver, were still alive to see the further fruits of her love and care. They both set me on those important first steps in mathematics, and for this I am eternally grateful.

Minneapolis
May 1998

Chapter 1

Prelude — Quadratic Polynomials and Quadratic Forms

Classical invariant theory is the study of the intrinsic properties of polynomials. By "intrinsic", we mean those properties which are unaffected by a change of variables and are hence purely geometric, untied to the explicit coordinate system in use at the time. Thus, properties such as factorizability and multiplicities of roots, as well as their geometrical configurations, are intrinsic, whereas the explicit values of the roots and the particular coefficients of the polynomial are not. The study of invariants is closely tied to the problem of equivalence — when can one polynomial be transformed into another by a suitable change of coordinates — and the associated canonical form problem — to find a system of coordinates in which the polynomial takes on a particularly simple form. The solution to these intimately related problems, and much more, are governed by the invariants, and so the first goal of classical invariant theory is to determine the fundamental invariants. With a sufficient number of invariants in hand, one can effectively solve the equivalence, and canonical form problems, and, at least in principle, completely characterize the underlying geometry of a given polynomial.

All of these issues are already apparent in the very simplest example — that of a quadratic polynomial in a single variable. This case served as the original catalyst for Boole and Cayley's pioneering work in the subject, [**24**, **36**], and can be effectively used as a simple (i.e., just high school algebra is required) concrete example that will motivate our study of the subject. We shall devote this introductory chapter to the elementary theory of quadratic polynomials in a single variable, together with homogeneous quadratic forms in two variables. Readers who are unimpressed with such relative trivialities are advised to proceed directly to the true beginning of our text in Chapter 2.

Quadratic Polynomials

Consider a quadratic polynomial in a single variable p:

$$Q(p) = ap^2 + 2bp + c. \tag{1.1}$$

Before addressing the question of what constitutes an invariant in this context, we begin our analysis with the elementary problem of determining a canonical form for the polynomial Q. In other words, we are trying to make Q as simple as possible by use of a suitable change of variable. As long as $a \neq 0$, the two roots of Q are, of course, given by the justly famous *quadratic formula*

$$p_+ = \frac{-b + \sqrt{-\Delta}}{a}, \qquad p_- = \frac{-b - \sqrt{-\Delta}}{a}, \tag{1.2}$$

where

$$\Delta = ac - b^2 \tag{1.3}$$

is the familiar *discriminant*[†] of Q. The existence of the two roots implies that we can factor

$$Q(p) = a(p - p_+)(p - p_-) \tag{1.4}$$

into two linear, possibly complex-valued, factors.

At this point, we need to be a bit more specific as to whether we are dealing with real or complex polynomials. Let us first concentrate on the slightly simpler complex version. The most obvious changes of variables preserving the class of quadratic polynomials are the affine transformations

$$\bar{p} = \alpha p + \beta, \tag{1.5}$$

for complex constants $\alpha \neq 0$ and β. Here α represents a (complex) scaling transformation,[‡] and β a complex translation. The transformation (1.5) maps the original quadratic polynomial $Q(p)$ to a new quadratic polynomial $\overline{Q}(\bar{p})$, which is constructed so that

$$\overline{Q}(\bar{p}) = \overline{Q}(\alpha p + \beta) = Q(p). \tag{1.6}$$

In particular, if p_0 is a root of $Q(p)$, then $\bar{p}_0 = \alpha p_0 + \beta$ will be a root of $\overline{Q}(\bar{p})$. For example, if $Q(p) = p^2 - 1$, and we apply the transformation

---

[†] The sign chosen for the discriminant is in accordance with later generalizations.

[‡] If we write $\alpha = re^{i\theta}$, then the modulus r will act by scaling, whereas the exponential $e^{i\theta}$ will induce a rotation in the complex p-plane; see p. 46.

$\bar{p} = 2p - 1$, then $\overline{Q}(\bar{p}) = \frac{1}{4}\bar{p}^2 + \frac{1}{2}\bar{p} - \frac{3}{4}$. The roots $p_\pm = \pm 1$ of Q are mapped to the roots $\bar{p}_+ = 1$ and $\bar{p}_- = -3$ of $\overline{Q}$.

For a general complex quadratic polynomial, there are only two cases to consider. If its discriminant is nonzero, $\Delta \neq 0$, then the roots of Q are distinct. We can translate one root, say, p_-, to be zero and then scale so that the second root takes the value 1. Thus, by a suitable choice of α and β we can arrange that $\overline{Q}$ has its roots at 0 and 1. Consequently, under complex affine transformations (1.5), every quadratic polynomial with distinct roots can be placed in the canonical form $\overline{Q}(\bar{p}) = k(\bar{p}^2 - \bar{p})$ for some $k \in \mathbb{C}$.

Exercise 1.1. Find the explicit formulas for α, β that will reduce a quadratic polynomial Q to its canonical form. Is the residual coefficient k uniquely determined? Determine the formula(e) for k in terms of the original coefficients of Q.

Exercise 1.2. An alternative canonical form for such quadratics is $\widetilde{Q}(\tilde{p}\,) = \tilde{k}\,(\tilde{p}^{\,2} + 1)$. Do the same exercise for this canonical form, and describe what is happening to the roots of Q.

On the other hand, if the discriminant of Q vanishes, so $ac = b^2$, then Q has a single double root p_0 and so factors as a perfect square: $Q(p) = a(p - p_0)^2$. Clearly this property is intrinsic — it cannot be altered by any change of coordinates. We can translate the double root to the origin, reducing Q to a multiple of the polynomial $\tilde{p}^{\,2}$, and then scale the coordinate $\tilde{p}$ to reduce the multiple to unity, leading to a canonical form, $\overline{Q} = \bar{p}^2$, for a quadratic polynomial with a double root.

We are not quite finished, since we began by assuming that the leading coefficient $a \neq 0$. If $a = 0$, but $b \neq 0$, then Q reduces to a linear polynomial with a single root, $p_0 = -c/b$. We can, as in the preceding case, translate this root to 0 and then scale, producing the canonical form $\overline{Q} = \bar{p}$ in this case. If $b = 0$ also, then Q is a constant, and, from the viewpoint of affine transformations (1.5), there is nothing that can be done. Thus, we have constructed a complete list of canonical forms for quadratic polynomials, under complex affine changes of coordinates. Note particularly that the discriminant Δ and the leading coefficient a play distinguished roles in the classification.

Exercise 1.3. Suppose Q and $\overline{Q}$ are related by an affine change of variables (1.5). Determine how their discriminants and leading coefficients are related.

Affine Canonical Forms for Complex Quadratic Polynomials			
I.	$k(p^2+1)$	$\Delta \neq 0,\ a \neq 0$	distinct roots
II.	p^2	$\Delta = 0,\ a \neq 0$	double root
III.	p	$a = 0,\ b \neq 0$	linear
IV.	c	$a = b = 0$	constant

The case of real polynomials under real affine changes of coordinates is similar, but there are a few more cases to consider. First, note that the roots (1.2) of a real quadratic polynomial are either both real or form a complex conjugate pair, depending on the sign of the discriminant. If Q has complex conjugate roots, meaning that its discriminant is positive, then it can never be mapped, under a real change of variables, to a quadratic polynomial with real roots, and so our complex canonical form is not as universally valid in this case. However, if the two roots are $p_\pm = r \pm is$, then a translation by $\beta = -r$ will move them onto the imaginary axis; this may be followed by a scaling to place them at $\pm i$. Thus, the canonical form in this case is $k(p^2+1)$. On the other hand, if the discriminant is negative, then Q has two distinct real roots, which can be moved to ± 1, leading to the alternative canonical form $k(p^2-1)$. The remaining cases are as in the complex version, since a double root of a real quadratic polynomial is necessarily real. We therefore deduce the corresponding table of real canonical forms.

Affine Canonical Forms for Real Quadratic Polynomials			
Ia.	$k(p^2+1)$	$\Delta > 0,\ a \neq 0$	complex conjugate roots
Ib.	$k(p^2-1)$	$\Delta < 0,\ a \neq 0$	distinct real roots
II.	p^2	$\Delta = 0,\ a \neq 0$	double root
III.	p	$a = 0,\ b \neq 0$	single root
IV.	c	$a = b = 0$	constant

Exercise 1.4. Determine the possible canonical forms for a complex cubic polynomial $Q(p) = ap^3 + bp^2 + cp + d$ under affine changes of coordinates. *Hint*: What are the possible root configurations?

Quadratic Forms and Projective Transformations

While affine changes of coordinates are immediately evident, they do not form the most general class that preserves the space of polynomials. In order to motivate a further extension, we begin by explaining the connection between homogeneous and inhomogeneous polynomials. Instead of the inhomogeneous polynomial (1.1) in a single variable, we consider the homogeneous quadratic polynomial

$$Q(x,y) = ax^2 + 2bxy + cy^2, \tag{1.7}$$

in two variables x, y, known classically as a *quadratic form*. Clearly we can recover the inhomogeneous quadratic polynomial $Q(p)$ from the associated quadratic form $Q(x,y)$ by setting $p = x$ and $y = 1$, so that $Q(p) = Q(p,1)$. On the other hand, the homogeneous version (1.7) can be directly constructed from $Q(p)$ according to the basic formula

$$Q(x,y) = y^2\, Q\left(\frac{x}{y}\right). \tag{1.8}$$

An affine change of coordinates (1.5) will induce a linear transformation mapping the quadratic form $Q(x,y)$ associated with $Q(p)$ to the quadratic form $\overline{Q}(\bar{x},\bar{y})$ associated with $\overline{Q}(\bar{p})$, as defined in (1.6). Clearly, the upper triangular linear transformation

$$\bar{x} = \alpha x + \beta y, \qquad \bar{y} = y, \qquad \alpha \neq 0, \tag{1.9}$$

will have the desired effect on the quadratic forms:

$$\overline{Q}(\bar{x},\bar{y}) = \overline{Q}(\alpha x + \beta y, y) = Q(x,y).$$

We conclude that the theory of inhomogeneous quadratic polynomials under affine coordinate changes is isomorphic to the theory of quadratic forms under linear transformations of the form (1.9).

Now, a crucial observation is that the class of quadratic forms is preserved under a much wider collection of transformation rules. Namely, *any* invertible linear change of variables

$$\bar{x} = \alpha x + \beta y, \qquad \bar{y} = \gamma x + \delta y, \qquad \alpha\delta - \beta\gamma \neq 0, \tag{1.10}$$

will map a homogeneous polynomial in x and y to a homogeneous polynomial in $\bar{x}$ and $\bar{y}$ according to

$$\overline{Q}(\bar{x},\bar{y}) = \overline{Q}(\alpha x + \beta y, \gamma x + \delta y) = Q(x,y). \tag{1.11}$$

The coefficients of the transformed polynomial

$$\overline{Q}(\bar{x},\bar{y}) = \bar{a}\,\bar{x}^2 + 2\bar{b}\,\bar{x}\bar{y} + \bar{c}\,\bar{y}^2$$

are constructed from those of the original polynomial (1.7) according to the explicit formulae

$$\begin{aligned} a &= \alpha^2\,\bar{a} + 2\alpha\gamma\,\bar{b} + \gamma^2\,\bar{c}, \\ b &= \alpha\beta\,\bar{a} + (\alpha\delta + \beta\gamma)\,\bar{b} + \gamma\delta\,\bar{c}, \\ c &= \beta^2\,\bar{a} + 2\beta\delta\,\bar{b} + \delta^2\,\bar{c}. \end{aligned} \tag{1.12}$$

Remarkably, the discriminant of the transformed polynomial is directly related to that of the original quadratic form — a straightforward computation verifies that they agree up to the square of the determinant of the coefficient matrix $A = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ for the linear transformation (1.10):

$$\Delta = ac - b^2 = (\alpha\delta - \beta\gamma)^2\,(\bar{a}\bar{c} - \bar{b}^2) = (\alpha\delta - \beta\gamma)^2\,\overline{\Delta}. \tag{1.13}$$

The transformation rule (1.13) expresses the underlying invariance of the discriminant of a quadratic polynomial and provides the simplest example of an invariant (in the sense of classical invariant theory).

Remark: A linear transformation (1.10) is called *unimodular* if it has unit determinant $\alpha\delta - \beta\gamma = 1$ and hence preserves planar areas. For the restricted class of unimodular transformations, the discriminant is a bona fide invariant: $\overline{\Delta} = \Delta$.

What is the effect of a general linear transformation on the original inhomogeneous polynomial? For this purpose, it helps to refer back to the formula (1.8) relating the inhomogeneous polynomial and its homogeneous counterpart. Specifically, the inhomogeneous or projective variable p is identified with the ratio of the homogeneous variables, so $p = x/y$. Therefore, the effect of the linear transformation (1.10) is to transform the projective variable p according to the *linear fractional* or *Möbius* or *projective transformation*

$$\bar{p} = \frac{\alpha p + \beta}{\gamma p + \delta}, \qquad \alpha\delta - \beta\gamma \neq 0. \tag{1.14}$$

Thus, each invertible 2×2 coefficient matrix $A = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ induces an invertible transformation mapping p to $\bar{p}$, which is defined everywhere except, when $\gamma \neq 0$, at the singular point $p = -\delta/\gamma$. Such transformations lie at the heart of projective geometry and are of fundamental importance, not just in invariant theory, but in a wide range of classical and modern disciplines, including complex analysis and geometry, [**8**], number theory, [**79**], and hyperbolic geometry, [**20**]. Indeed, this rather simple construction has led to some of the most profound consequences in all of mathematics.

Exercise 1.5. Prove directly that the composition of two linear fractional transformations is again a linear fractional transformation, whose coefficients are obtained by multiplying the associated 2×2 coefficient matrices. In particular, the inverse of a linear fractional transformation is the linear fractional transformation determined by the inverse coefficient matrix.

Exercise 1.6. Show that two coefficient matrices A and $\widetilde{A}$ determine the same linear fractional transformation (1.14) if and only if they are scalar multiples of each other: $A = \lambda \widetilde{A}$. Thus, in the complex case, any linear fractional transformation (1.14) can be implemented by a unimodular coefficient matrix: $\det A = 1$. What is the unimodular linear transformation associated with the affine transformation (1.5)? Is this result valid in the real case?

How should a general linear fractional transformation act on an inhomogeneous quadratic polynomial (1.1)? We want to maintain the transformation rules (1.12) on the coefficients, so that the action will be the inhomogeneous counterpart to the linear action (1.11) on homogeneous quadratic forms. This requires that the quadratic polynomials $Q(p)$ and $\overline{Q}(\bar{p})$ are related according to the basic formula

$$Q(p) = (\gamma p + \delta)^2 \, \overline{Q}(\bar{p}) = (\gamma p + \delta)^2 \, \overline{Q}\left(\frac{\alpha p + \beta}{\gamma p + \delta}\right). \tag{1.15}$$

Note that the additional factor $(\gamma p + \delta)^2$, known as the quadratic *multiplier*, is used to clear denominators so that the linear fractional transformation (1.14) will still map quadratic forms to quadratic forms. The reader might enjoy verifying that the transformation rules (1.15) does lead to exactly the same formulae (1.12) for the coefficients, and hence the discriminant continues to satisfy the basic invariance criterion (1.13). Note that, even though two coefficient matrices which are scalar multiples of each other determine the *same* linear fractional transformation (1.14), their action on quadratic polynomials (1.15) is *different* (unless $A = \pm \widetilde{A}$), owing to the effect of the multiplier.

Exercise 1.7. Show that the inversion $\bar{p} = 1/(p+1)$ maps the quadratic polynomial $Q(p) = p^2 - 1$ to the linear polynomial $\overline{Q}(\bar{p}) = -2\bar{p} + 1$. Thus projective transformations do not necessarily preserve the degree of a polynomial. Given a linear fractional transformation (1.14), determine which quadratic polynomials $Q(p)$ are mapped to linear polynomials. Which are mapped to constant polynomials?

Let us return to the canonical form problem for quadratic polynomials, now rearmed with the more general projective transformations. Clearly, by suitably combining the transformations and appealing to Exercise 1.5, we can begin by placing the quadratic polynomial in canonical affine form. Consider first the complex canonical form $Q = k(p^2 + 1)$. If we scale according to the coefficient matrix $A = \lambda \mathbb{1}$, where $\mathbb{1}$ is the 2×2 identity matrix and $\lambda^2 = k$, then we can normalize $Q \mapsto p^2 + 1$. Furthermore, the transformation $\bar{p} = (p - i)/(p + i)$ will map $p^2 + 1$ to the linear polynomial $\bar{p}$. Therefore, if $\Delta \neq 0$, and so $Q(p)$ either has two distinct roots or is a nonzero linear polynomial, then there is just one canonical form, namely $Q(p) = p$. On the other hand, if we take the affine canonical form $Q = p^2$, we can apply the inversion $\bar{p} = 1/p$ to map it to the constant polynomial $\overline{Q} = 1$; further, any other constant (nonzero) polynomial can, by applying a diagonal scaling matrix, be mapped to the constant 1. Thus, for complex quadratic polynomials under general linear fractional transformations, there are only three canonical forms: the first is p, which occurs when $\Delta \neq 0$; the second is 1, which occurs when $\Delta = 0$ but Q is not identically 0; and the last is the most trivial case, namely $Q \equiv 0$.

Canonical Forms for Complex Quadratic Polynomials

I.	p	$\Delta \neq 0$	distinct roots
II.	1	$\Delta = 0$, $Q \not\equiv 0$	double root
III.	0	$Q \equiv 0$	

Thus, under projective transformations, every complex quadratic polynomial is equivalent to a linear or constant polynomial. Since the action of linear fractional transformations on inhomogeneous quadratic polynomials mirrors that of linear transformations on homogeneous quadratic forms, each of our canonical forms has a homogeneous counterpart. We conclude that, under complex linear transformations, there are also three different canonical quadratic forms: first, xy, or, alternatively, $x^2 + y^2$; second, x^2; and, third, the trivial zero form 0.

In the real case, note that the transformation rules (1.3) for the discriminant imply that its sign is invariant: if $\Delta > 0$, say, then $\overline{\Delta} > 0$ also. Of course, this just means that one cannot map real roots to complex roots by a real projective transformation. Moreover, the sign of Q itself is

also invariant; one cannot map a positive definite quadratic polynomial to an indefinite or negative definite one. Consequently, there are three different canonical forms with nonvanishing discriminant. The sign of Q also affects the classification of quadratics with vanishing discriminant.

Canonical Forms for Real Quadratic Polynomials

Ia.	$p^2 + 1$	$\Delta > 0,\ Q \geq 0$	complex roots
Ib.	$-p^2 - 1$	$\Delta > 0,\ Q \leq 0$	complex roots
Ic.	p	$\Delta < 0$	distinct real roots
IIa.	1	$\Delta = 0,\ Q \geq 0$	double root
IIb.	-1	$\Delta = 0,\ Q \leq 0$	double root
III.	0	$Q \equiv 0$	

The corresponding homogeneous canonical forms are the positive definite, $x^2 + y^2$, negative definite, $-x^2 - y^2$, and indefinite, which can be taken as either xy or $x^2 - y^2$, all of which were complex-equivalent, followed by the degenerate cases x^2, $-x^2$, and 0.

Suppose we restrict to area-preserving transformations, with unimodular coefficient matrix: $\det A = 1$. In this case, the discriminant is strictly invariant, and hence we can no longer rescale to normalize it to be ± 1. Retracing the preceding arguments, we see that the only effect is to introduce an extra scaling factor into the list of canonical forms. Thus, for complex-valued quadratic polynomials under area-preserving changes of variables, the canonical forms having nonzero discriminant become a one-parameter family of linear forms kp. Note that the inversion $\tilde{p} = -1/p$ will map kp to $-k\tilde{p}$, both of which have discriminant $\Delta = k^2$, but otherwise one cannot transform between two different linear canonical forms. Therefore, a complete list of canonical forms for complex quadratic polynomials under unimodular linear fractional transformations consists of the linear forms kp, along with the constant forms 1 and 0. In the real case, one similarly finds two families of canonical forms, $k(p^2 + 1)$ and kp, which are distinguished by the sign of the discriminant. In the degenerate cases where $\Delta = 0$, the list of canonical forms remains the same as before.

Remark: A generic unimodular linear fractional transformation depends on three free parameters: α, β, and γ. Further, a quadratic polynomial (1.1) has three coefficients. Thus, one might expect that one could normalize all three coefficients via a suitable choice of the three parameters in the linear fractional transformation. The invariance of the discriminant proves that this naïve parameter count can be misleading. (See Chapter 8 for a more sophisticated and accurate version, which is based on the orbit dimensions.)

Exercise 1.8. Determine the canonical forms for complex-valued quadratic polynomials under the class of real linear (or linear fractional) transformations. In other words, the coefficients a, b, c in (1.7) or (1.1) are allowed to be complex, but the transformations (1.10) or (1.14) are restricted so that $\alpha, \beta, \gamma, \delta$ are all real.

This concludes our brief presentation of the admittedly elementary theory of quadratic polynomials in one complex or one real variable. Extensions to multi-dimensional quadratic forms are certainly of interest, and we shall briefly return to this topic in Chapters 3 and 10. However, our more immediate interest is in extending these basic considerations to higher degree polynomials in a single variable and/or homogeneous polynomials in two variables. In the classical literature, these are known as "binary forms". Their invariants, geometry, and canonical forms, under projective and/or linear transformations, constitute the heart and soul of the classical theory.

Chapter 2

Basic Invariant Theory for Binary Forms

Using the previous chapter as our motivational springboard, let us now dive into our chosen subject. Most of the classical literature, and indeed most of the present text, is devoted to the simplest case — that of a binary form or homogeneous polynomial in two variables, along with the inhomogeneous univariate counterpart. In this chapter, we shall introduce many of the fundamental concepts in the invariant theory of binary forms. The ideas will be illustrated by the next two most important cases — that of cubic and quartic polynomials. In each case, we shall exhibit a complete system of invariants, as well as a complete list of canonical forms. These examples will serve to motivate the general definitions of invariants and covariants. The emphasis here is on important particular examples, such as Hessians, resultants, and discriminants, and their role in the classification and geometry of binary forms. These initial constructions bring the basic problem of classifying the invariants into focus, leading to the fundamental Basis Theorem of Hilbert, whose proof will appear in Chapter 9. The chapter concludes with a brief discussion of the algebraic relationships, known as "syzygies", that exist among the fundamental invariants and covariants.

Binary Forms

In the classical literature, homogeneous polynomials are called *forms*.† The adjectives "binary", "ternary", and so on refer to the number of variables that the form depends on. The most important case, and the one we shall primarily concentrate on, is that of a *binary form*

$$Q(\mathbf{x}) = Q(x, y) = \sum_{i=0}^{n} \binom{n}{i} a_i \, x^i y^{n-i}, \tag{2.1}$$

† The classical term "form" (which replaced Cayley's older "quantic", [**40**]) as used here should not be confused with the modern term "differential form". In this book, all forms are symmetric (and hence polynomials) — as opposed to the anti-symmetric forms of importance in geometry and topology, [**25**, **168**].

which is a homogeneous polynomial function of two variables $\mathbf{x} = (x, y)$. (The binomial coefficients $\binom{n}{i}$ are introduced for later convenience.) We shall consider both real and complex† forms, as the methods apply equally well to both. The number $n \in \mathbb{N}$ is the *degree* of the form, and we note that Q satisfies the basic homogeneity equation $Q(\lambda \mathbf{x}) = \lambda^n Q(\mathbf{x})$.

As with the correspondence between quadratic polynomials and quadratic forms, each homogeneous polynomial (2.1) will correspond to an inhomogeneous polynomial

$$Q(p) \equiv Q(p, 1) = \sum_{i=0}^{n} \binom{n}{i} a_i p^i, \tag{2.2}$$

depending on a single scalar variable p. At the risk of initial confusion, we shall use the same symbol Q for both the homogeneous form (2.1) as well as its inhomogeneous counterpart (2.2). The reader might wish to insert some extra notation, such as $\widetilde{Q}(p)$, to distinguish the inhomogeneous version (2.2), because two different homogeneous forms might, ostensibly, reduce to the "same" inhomogeneous counterpart. For example, the linear form $Q_1(x, y) = x + 2y$ has inhomogeneous version $\widetilde{Q}_1(p) = p + 2$; the quadratic form $Q_2(x, y) = xy + 2y^2$ also has $\widetilde{Q}_2(p) = p + 2$. However, this identification is, in fact, illusory; the former is a linear (rather affine) polynomial, whereas the latter should be regarded not as a linear polynomial, but rather as a degenerate quadratic polynomial! The distinction is, at the outset, certainly not evident, but it will become so once the transformation rules are brought into play. In point of fact, the use of a notation like $\widetilde{Q}$ tends, I believe, to be *more* confusing than our agreement to use the same notation for homogeneous and inhomogeneous polynomials. In the preceding example, then, we would have $Q_1(p) = p + 2$, and $Q_2(p) = p + 2$, but $Q_1 \neq Q_2$ because they come from different homogeneous representatives! Any reader who is willing to persevere should soon recover from this initial confusion.

Remark: Actually, the difficulty we are experiencing at this juncture is reflective of the fact that the inhomogeneous representative of a homogeneous function is not really a function at all, but, rather, a section of a "line bundle" over a one-dimensional base space, cf. [**25**, **161**]. Thus, the fact that $Q_1 \neq Q_2$ even though they both have the same co-

† Polynomials whose coefficients belong to other fields are certainly of great algebraic interest, but such extensions would take us too far "afield".

ordinate formula, is because Q_1 is a section of the "linear line bundle", whereas Q_2 is a section of the "quadratic line bundle", and we just happen to have chosen the underlying coordinate p so that they have the same formula. But to keep the exposition reasonably elementary, I have chosen not to adopt this more advanced geometric framework.

Given an inhomogeneous polynomial $Q(p)$, we can recover its homogeneous form $Q(x, y)$ via the simple rule

$$Q(x, y) = y^n \, Q\left(\frac{x}{y}\right), \tag{2.3}$$

provided we specify its degree n in advance. Formula (2.3) proves that there is a one-to-one correspondence $Q(x, y) \iff (Q(p), n)$ between homogeneous forms and inhomogeneous polynomials once we append the latter's degree. The degree n of the form might be larger than the naïve degree of $Q(p)$, meaning the degree of its leading term. Note that the naïve degree of $Q(p)$ will be strictly less than the degree of (2.2) if and only if its leading coefficient vanishes: $a_n = 0$. In classical invariant theory, the naïve degree is more or less meaningless since it can be changed by a suitable transformation, whereas the true degree is intrinsic and extremely important.

Definition 2.1. The *degree* of a (nonzero) homogeneous form $Q(x, y)$ is the degree of any of its terms. The *degree* of an inhomogeneous polynomial $Q(p)$ is the degree of its homogeneous representative. Two inhomogeneous polynomials are considered to be equal if and only if they have the same coordinate formula *and* the same degree.

Transformation Rules

Classical invariant theory is concerned with the intrinsic geometric properties of forms, meaning those properties which do not depend on the introduction of a particular coordinate system. In the case of homogeneous forms (2.1), we are naturally led to consider the effect of invertible linear changes of variables

$$\bar{x} = \alpha x + \beta y, \qquad \bar{y} = \gamma x + \delta y, \qquad \alpha\delta - \beta\gamma \neq 0, \tag{2.4}$$

in which the coefficient matrix $A = \left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ is nonsingular, and either real or complex, depending on the type of form under consideration. Note that (2.4) defines (essentially) the most general class of transformations on a two-dimensional space which preserve the class of homogeneous

polynomials of a fixed degree.† Under such a linear transformation, the polynomial $Q(x, y)$ is mapped to a new polynomial $\overline{Q}(\bar{x}, \bar{y})$, defined so that

$$\overline{Q}(\bar{x}, \bar{y}) = \overline{Q}(\alpha x + \beta y, \gamma x + \delta y) = Q(x, y). \tag{2.5}$$

Thus, the matrix A induces a transformation on the coefficients a_i of Q, mapping them into the corresponding coefficients $\bar{a}_i$ of $\overline{Q}$. It is not difficult to determine precise formulae for the coefficients of the transformed polynomial.

Theorem 2.2. *Let $Q(x, y)$ and $\overline{Q}(\bar{x}, \bar{y})$ be two binary forms related by the transformation rule* (2.5). *Then their coefficients are related by the explicit formulae*

$$a_i = \sum_{k=0}^{n} \bar{a}_k \left\{ \sum_{j=\max\{0, i+k-n\}}^{\min\{i,k\}} \binom{i}{j}\binom{n-i}{k-j} \alpha^j \beta^{k-j} \gamma^{i-j} \delta^{n+j-i-k} \right\}, \qquad i = 0, \dots, n. \tag{2.6}$$

Note that (2.6) reduces to the quadratic transformation formulae (1.12) when $n = 2$. Theorem 2.2 is a straightforward consequence of the Binomial Theorem. In essence, classical invariant theory for binary forms can be regarded as the analysis of the consequences of these specific transformation rules. However, we shall make surprisingly little use of the complicated explicit formulae (2.6); all of the major techniques can be developed without any direct reference to them.

The induced action of a linear transformation (2.4) on the projective coordinate $p = x/y$ is, as in the case of quadratic forms, governed by linear fractional transformations

$$\bar{p} = \frac{\alpha p + \beta}{\gamma p + \delta}, \qquad \alpha\delta - \beta\gamma \neq 0. \tag{2.7}$$

† However, specific polynomials may admit more general types of transformations which preserve their underlying form. Also, any invertible homogeneous polynomial transformation will map a homogeneous polynomial to another homogeneous polynomial, albeit of a different degree. It is outside the scope of this book to consider this more general class, which includes the Tschirnhaus transformations, [**42**], [**58**; p. 210], that are classically used to reduce higher degree polynomials to canonical form. Even the classification of invertible polynomial transformations remains rather rudimentary. For instance, the classical "Jacobian conjecture" — a polynomial transformation with constant Jacobian determinant is invertible — remains unsolved, [**18**].

The transformation rule for inhomogeneous polynomials is a simple consequence of the basic correspondence (2.3).

Proposition 2.3. *Let $Q(x, y)$ and $\overline{Q}(\bar{x}, \bar{y})$ be homogeneous polynomials of the same degree n which are related by a linear change of variables according to* (2.5). *Then the associated inhomogeneous polynomials $Q(p)$ and $\overline{Q}(\bar{p})$ are related by the basic linear fractional transformation rule of degree n:*

$$Q(p) = (\gamma p + \delta)^n \, \overline{Q}(\bar{p}) = (\gamma p + \delta)^n \, \overline{Q}\left(\frac{\alpha p + \beta}{\gamma p + \delta}\right). \tag{2.8}$$

As before, the role of the n^{th} order *multiplier* $(\gamma p + \delta)^n$ is to clear denominators so that the linear fractional transformation (2.7) will map polynomials to polynomials of the same degree. It is easy to see that the coefficients a_i of the inhomogeneous form $Q(p)$ are subjected to the *same* transformation rules (2.6) under the linear fractional transformation rule (2.8) as those of the homogeneous representative $Q(x, y)$. Thus, the degree of an inhomogeneous polynomial is not necessarily specified by its local coordinate formula (2.2) but rather is distinguished by its behavior under linear fractional changes of coordinates.

The Geometry of Projective Space

Before proceeding further, it will help if we review elementary projective geometry, which underlies the correspondence between homogeneous forms and their inhomogeneous counterparts. Projective geometry dates back (at least) to the Renaissance, when European artists developed perspective representations of scenes. The mathematical foundations begin with the work of Desargues and Pascal, and were brought to maturity by Poncelet; see [**26**] for historical details. This subject was one of the mainstays of classical mathematics, while recent advances in image processing and computer vision, cf. [**157**], have underscored its continued relevance.

Definition 2.4. Given a vector space V, the associated *projective space* $\mathbb{P}(V)$ is defined as the set of all one-dimensional subspaces of V, i.e., the set of all lines through the origin in V.

If V is finite-dimensional, then its projective space forms a "manifold"† whose dimension is one less than that of V itself. The simplest

† See Chapter 8 for the precise definition.

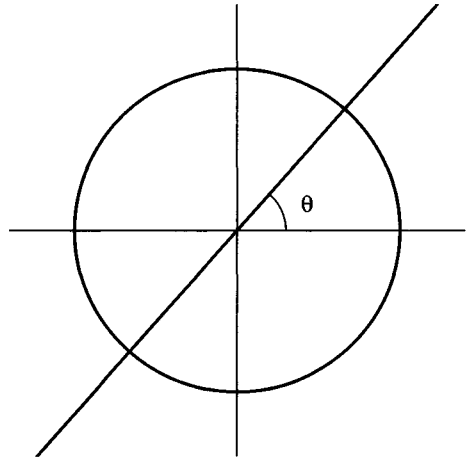

Figure 1. The Real Projective Line.

example is the real projective line, $\mathbb{RP}^1 = \mathbb{P}(\mathbb{R}^2)$, which is the projective space associated with the real plane, i.e., the space of lines through the origin in $\mathbb{R}^2$. Each line intersects the unit circle $S^1 \subset \mathbb{R}^2$ twice — see Figure 1 — and thus we can identify $\mathbb{RP}^1$ with the circle obtained by identifying opposite (antipodal) points on S^1. (To see that this identification does, in fact, produce a circle, we note that the map $\theta \mapsto 2\theta$ from S^1 to itself will identify the antipodal points in a unique manner.) Thus, the angle $0 \leq \theta < \pi$ that each line makes with the horizontal can be used to coordinatize the real projective line.

Classically, one views the Cartesian coordinates on $\mathbb{R}^2$ as defining *homogeneous coordinates* on $\mathbb{RP}^1$ and employs a square bracket to indicate this fact. Thus, a nonzero coordinate pair $0 \neq (x, y) \in \mathbb{R}^2$ defines the homogeneous coordinate $[\,x, y\,]$ of the line passing through it. Homogeneous coordinates are defined only up to scalar multiple, so that $[\,\lambda x, \lambda y\,] = [\,x, y\,]$ for any $\lambda \neq 0$. If a line is not horizontal, then it intersects the line $y = 1$ at a unique point $[\,p, 1\,]$. Therefore, instead of the angle θ the line makes with the horizontal, we may adopt the horizontal component $p = \cot\theta$ of its intersection with the line $y = 1$ as our preferred coordinate; see Figure 2. If $[\,x, y\,]$ is any other homogeneous coordinate for the given line, then its canonical representative $[\,p, 1\,] = [\,x/y, 1\,]$ is obtained by multiplying by the scalar $\lambda = 1/y$. Thus, the open subset of $\mathbb{RP}^1$ consisting of the non-horizontal lines can be identified with $\mathbb{R}$ itself, with $p = x/y$ providing the *projective coordinate*. In

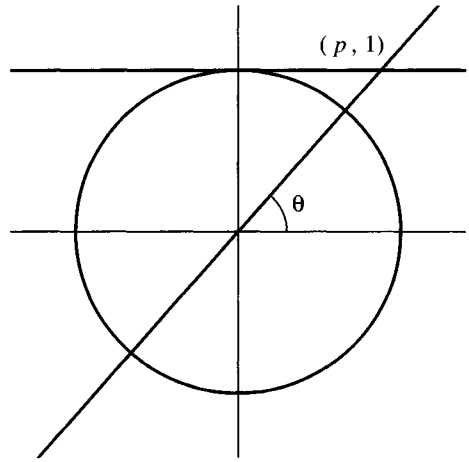

Figure 2. The Projective Coordinate.

this way, we can regard the real projective line as the "completion" of the ordinary real line by adjoining a single point at infinity, which corresponds to the horizontal line through the origin. Alternatively, we can omit the vertical line and use the canonical coordinate $[1, q]$, where $q = y/x = \tan\theta$, to represent a different open subset of $\mathbb{RP}^1$. The change of coordinates from the non-horizontal to the non-vertical cases is the inversion $q = 1/p$.

A linear transformation (1.10) on $\mathbb{R}^2$ will induce a linear fractional transformation (2.7) on $\mathbb{RP}^1$. This is because the line with homogeneous coordinates $[p, 1]$ is mapped to the line with homogeneous coordinates $[\alpha p+\beta, \gamma p+\delta]$, whose canonical representative is (assuming $\gamma p+\delta \neq 0$) given by $[(\alpha p+\beta)/(\gamma p+\delta), 1]$. Moreover, the projective transformation (2.7) remains globally defined on $\mathbb{RP}^1$ — the point $p = -\delta/\gamma$ is mapped to the point at ∞ (indicating that the line through $(-\delta, \gamma)$ is mapped to the horizontal line), whereas the point at ∞ is mapped to the point α/γ. (If $\gamma = 0$, then the point at ∞ stays there, since such maps fix the horizontal line.) Note that the scalings $(x, y) \mapsto (\lambda x, \lambda y)$, corresponding to scalar multiples of the identity matrix, have trivial action on $\mathbb{RP}^1$ since they fix every line that passes through the origin.

Similar considerations apply to the complex projective line $\mathbb{CP}^1$, which is the projective space $\mathbb{P}(\mathbb{C}^2)$ corresponding to the two-dimensional complex linear space $\mathbb{C}^2$. One can identify $\mathbb{CP}^1$ with the usual Riemann sphere S^2 of complex analysis, [**8**], which can be viewed as the

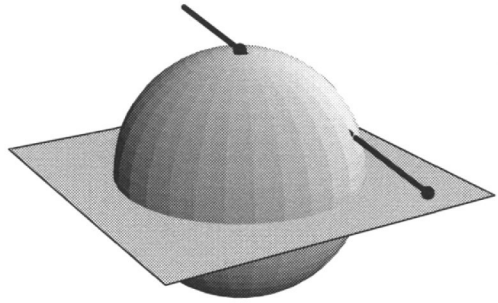

Figure 3. Stereographic Projection.

completion of the complex line† $\mathbb{C}$ by adjoining a point at ∞. Explicitly, given $0 \neq (z, w) \in \mathbb{C}^2$, we define the real variables ξ, η, ζ by

$$\xi + i\eta = \frac{2\, z\, \overline{w}}{|\, z\,|^2 + |\, w\,|^2}\,, \qquad \zeta = \frac{|\, z\,|^2 - |\, w\,|^2}{|\, z\,|^2 + |\, w\,|^2}\,. \tag{2.9}$$

The reader can check that the point (ξ, η, ζ) necessarily lies on the unit sphere $S^2 \subset \mathbb{R}^3$, that is, $\xi^2 + \eta^2 + \zeta^2 = 1$. Moreover, two points (z, w) and (z', w') in $\mathbb{C}^2$ map to the same point $(\xi, \eta, \zeta) \in S^2$ if and only if they are complex scalar multiples of each other, so that $z' = \lambda z$, $w' = \lambda w$, for some $0 \neq \lambda \in \mathbb{C}$.

Exercise 2.5. The usual method for mapping the unit sphere $\xi^2+\eta^2+\zeta^2 = 1$ to the x, y coordinate plane is by stereographic projection from the north pole $(0, 0, 1)$. Geometrically, the stereographic image of a point $(\xi, \eta, \zeta) \in S^2$ which is *not* the north pole is the point $(u, v, 0)$ obtained by intersecting the line connecting (ξ, η, ζ) to the north pole with the plane — see Figure 3. Show that the stereographic image of a point is defined by the formulae

$$u = \frac{\xi}{1 - \zeta}\,, \qquad v = \frac{\eta}{1 - \zeta}\,. \tag{2.10}$$

Prove further that the map from $\mathbb{C}^2 \setminus \{0\}$ to $\mathbb{C} \simeq \mathbb{R}^2$ obtained by first

† Here is a potential source of confusion. In elementary mathematics, one speaks of the "complex plane" since the set of complex numbers $\mathbb{C}$ is visualized as a two-dimensional plane (and, indeed, is a *real* two-dimensional vector space); however, as a complex vector space, $\mathbb{C}$ is one-dimensional and will therefore be referred to as the "complex line", with $\mathbb{C}^2$ being the "genuine complex plane". To minimize misunderstanding, we shall try to avoid using the term "complex plane" in this book.

mapping to the Riemann sphere via (2.9) and then applying stereographic projection (2.10) is the same as the projective coordinate map $u + iv = z/w$, for $(z, w) \in \mathbb{C}^2$ with $w \neq 0$.

All of our constructions for the real projective line have their complex counterparts, which are found just by letting all quantities assume arbitrary complex values. Indeed, we shall usually use x and y rather than z and w to denote complex coordinates on $\mathbb{C}^2$, so that the real and complex algebraic formulae are identical. The subset of complex lines not parallel to the x-axis can be given the homogeneous coordinate $[p, 1]$, where $p = x/y$ is the projective coordinate, and can thus be identified with the ordinary complex line $\mathbb{C}$ with coordinate p. Complex linear changes of variables on $\mathbb{C}^2$, as in (2.4), induce complex linear fractional transformations on $\mathbb{CP}^1$, as in (2.7).

Homogeneous Functions and Forms

The correspondence between a vector space and its associated projective space induces a correspondence between homogeneous functions and their inhomogeneous counterparts, generalizing the correspondence between quadratic forms and ordinary quadratic polynomials. Let us illustrate this correspondence in the particular case of the real projective line. Let $\pi\colon \mathbb{R}^2 \setminus \{0\} \to \mathbb{RP}^1$ denote the map that takes a point in $\mathbb{R}^2$ to the line that connects it to the origin. In terms of our projective coordinates, $p = \pi(x, y) = x/y$ for points (x, y) not on the horizontal axis. Any real-valued function[†] $F\colon \mathbb{RP}^1 \to \mathbb{R}$ on the projective line induces a function $Q\colon \mathbb{R}^2 \to \mathbb{R}$, which is given by composition: $Q = F \circ \pi$. In other words, given $F(p)$, the corresponding homogeneous function is

$$Q(x, y) = F\left(\frac{x}{y}\right). \tag{2.11}$$

Clearly, though, formula (2.11) does not reproduce the correspondence (2.3) between homogeneous and inhomogeneous polynomials. For instance, if $F(p) = p^2 + 1$, then $Q(x, y) = y^{-2}x^2 + 1$, which is not even defined on the x-axis. The key point is that (2.11) defines a function Q which is homogeneous of degree zero, $Q(\lambda \mathbf{x}) = Q(\mathbf{x})$, and hence can never (unless Q is constant) define a homogeneous polynomial on $\mathbb{R}^2$.

[†] The mapping notation $F\colon X \to Y$ does not necessarily imply that F is defined everywhere on X.

Definition 2.6. A function $Q: \mathbb{R}^2 \to \mathbb{R}$ is called *homogeneous* of *degree* n if it satisfies the basic homogeneity equation

$$Q(\lambda \mathbf{x}) = \lambda^n Q(\mathbf{x}), \qquad \text{for all } \mathbf{x} \in \mathbb{R}^2. \tag{2.12}$$

In particular, homogeneous polynomials of degree n are homogeneous functions, of positive integral degree n; on the other hand, not every homogeneous function, even those of positive integral degree, is a polynomial. The following simple characterization of homogeneous functions is attributed to Euler.

Theorem 2.7. *A differentiable function Q is homogeneous of degree n if and only if it satisfies* Euler's formula, *which is the first order partial differential equation*

$$x \frac{\partial Q}{\partial x} + y \frac{\partial Q}{\partial y} = nQ. \tag{2.13}$$

Proof: Equation (2.13) follows directly from (2.12) by differentiating with respect to λ:

$$\frac{\partial}{\partial \lambda} Q(\lambda x, \lambda y) = x \frac{\partial Q}{\partial x}(\lambda x, \lambda y) + y \frac{\partial Q}{\partial y}(\lambda x, \lambda y) = n\lambda^{n-1} Q(x, y). \tag{2.14}$$

Setting $\lambda = 1$ yields (2.13). Conversely, if Q satisfies (2.13), then the first equality in (2.14) implies that

$$\frac{\partial}{\partial \lambda} Q(\lambda x, \lambda y) = \frac{n}{\lambda} Q(\lambda x, \lambda y). \tag{2.15}$$

Fixing x and y, we regard (2.15) as an ordinary differential equation for the function $h(\lambda) = Q(\lambda x, \lambda y)$, namely, $dh/d\lambda = (n/\lambda)\, h$. This equation can be readily integrated; the resulting solution $h(\lambda) = \lambda^n h(1) = \lambda^n Q(x, y)$ recovers the homogeneity condition (2.12). *Q.E.D.*

Exercise 2.8. Show that the function

$$Q(x, y) = \begin{cases} x^2 \exp(x^2/y^2), & y \neq 0, \\ 0, & y = 0, \end{cases}$$

is a smooth (meaning infinitely differentiable), globally defined homogeneous function of degree 2. Are there any analytic, globally defined homogeneous functions other than homogeneous polynomials?

A simple modification of the direct formula (2.11) will allow us to construct homogeneous functions of arbitrary degree from functions on the projective space. First, we remark that the product of a homogeneous function of degree m with a homogeneous function of degree n

is also a homogeneous function of degree $m + n$. Therefore, if $Q_0(\mathbf{x})$ is a particular nonzero homogeneous function of degree n, then any other homogeneous function of degree n can be written as a product $Q(\mathbf{x}) = Q_0(\mathbf{x})R(\mathbf{x})$, where $R(\mathbf{x})$ is an arbitrary homogeneous function of degree 0. In particular, choosing $Q_0(x, y) = y^n$ allows us to conclude the general version of the correspondence (2.3).

Proposition 2.9. *Every homogeneous function $Q(x, y)$ of degree n can be written in the form $Q(x, y) = y^n F(x/y)$, where $F(p)$ is an arbitrary function on $\mathbb{RP}^1$.*

As with polynomials, the feature that distinguishes the different homogeneous representatives of a given inhomogeneous function is how they behave under changes of variables. Consequently, we cannot speak of a function $F: \mathbb{RP}^1 \to \mathbb{R}$ on projective space *in vacuo*, since (*a*) it does not tell us what degree its homogeneous representative should be, and (*b*) it does not tell us how it behaves under changes of variables. Only when we specialize to homogeneous functions of a fixed degree are the correspondences and transformation rules unambiguous.

Roots

As we saw in Chapter 1, the roots of quadratic polynomials play a critical role in their classification. We expect the geometrical configurations of the roots of more general polynomials to play a similar role in their classification and the structure of their invariants. We begin with the basic definition.

Definition 2.10. Let $Q(p)$ be a function defined on the projective line. A *root* of Q is a point p_0 where Q vanishes: $Q(p_0) = 0$.

The key remark is that the concept of a root is independent of the coordinate system used to characterize the inhomogeneous function. Indeed, referring to the basic transformation rule (2.8), we see that, provided $\gamma p_0 + \delta \neq 0$, then $Q(p_0) = 0$ if and only if $\overline{Q}(\bar{p}_0) = 0$, where $\bar{p}_0 = (\alpha p_0 + \beta)/(\gamma p_0 + \delta)$ is the transformed root. On the other hand, if $\gamma p_0 + \delta = 0$, then the transformed root is at $\bar{p}_0 = \infty$, and so the coordinate formula breaks down. Nevertheless, the root still persists, and one says that the transformed function $\overline{Q}(\bar{p})$ has a root at ∞.

Each root p_0 of the inhomogeneous representative will correspond to an entire line of solutions to the homogeneous equation $Q(\mathbf{x}) = 0$. Indeed, (2.12) implies that if $\mathbf{x}_0 = (x_0, y_0)$ is a solution, so is any nonzero

scalar multiple $\lambda \mathbf{x}_0$, $\lambda \neq 0$. We will not distinguish between such solutions, since they all determine the same point in the projective space.

Definition 2.11. Let $Q(\mathbf{x})$ be a homogeneous function. By a *homogenized root* of Q we mean a line $\{\lambda \mathbf{x}_0\}$, $\mathbf{x}_0 \neq 0$, through the origin where (except possibly at the origin itself) Q vanishes.

Each homogenized root $\mathbf{x}_0 = [x_0, y_0]$ with $y_0 \neq 0$ corresponds to a root $p_0 = x_0/y_0$ of $Q(p)$. If $[x_0, 0]$ is a homogenized root, then it corresponds to the "infinite" root ∞ of the inhomogeneous form $Q(p)$. For instance, $Q(x, y) = xy + 2y^2$ has two homogenized roots: the lines through $(-2, 1)$ and $(1, 0)$; its inhomogeneous representative $Q(p) = p + 2$, which has degree 2, has roots at $p = -2$ and at $p = \infty$.

In the case of complex-valued polynomials, the Fundamental Theorem of Algebra tells us precisely how many roots there are. The key result is that every complex polynomial has at least one root.

Lemma 2.12. *Let $Q(p)$ be a nonconstant complex polynomial. Then there exists a point $p_0 \in \mathbb{C}$ such that $Q(p_0) = 0$.*

Many different proofs of this seminal result exist, and we refer the interested reader to [**71**], [**222**; Chapter 11] for details. Once we establish the existence of at least one complex root, then the polynomial admits a linear factor, and so the complete factorization of any complex polynomial follows by a straightforward induction.

Theorem 2.13. *Let*

$$Q(p) = c_m p^m + c_{m-1} p^{m-1} + \cdots + c_1 p + c_0 \tag{2.16}$$

be a polynomial with nonzero leading coefficient, $c_m \neq 0$. Then Q can be uniquely factored into a product of linear polynomials:

$$Q(p) = c_m \prod_{\nu=1}^{m} (p - p_\nu), \tag{2.17}$$

where $p_1, \ldots, p_m$ are the finite complex roots of Q.

Exercise 2.14. Prove that any real polynomial can be factored, over the reals, into a product of linear and quadratic factors. *Hint*: Use the fact that the complex roots of a real polynomial appear in complex conjugate pairs.

Definition 2.15. A root p_0 of a polynomial $Q(p)$ is said to have *multiplicity* k if we can write $Q(p) = (p - p_0)^k R(p)$, where $R(p)$ is a polynomial with $R(p_0) \neq 0$. In other words, p_0 is a root of of multiplicity

k if and only if the linear factor $p - p_0$ appears precisely k times in the factorization (2.17).

Exercise 2.16. Prove that p_0 is a root of $Q(p)$ of multiplicity k if and only if Q and its first $k-1$ derivatives vanish there: $Q(p_0) = Q'(p_0) = \cdots = Q^{(k-1)}(p_0) = 0$, but $Q^{(k)}(p_0) \neq 0$.

So far, we have been a bit cavalier with our presentation, since we have been ignoring the "true" degree of the polynomial $Q(p)$, meaning the degree of its homogeneous representative $Q(x, y)$, in lieu of its naïve degree, as determined by the degree of its leading term. Since the naïve degree of a polynomial can change under projective transformations, we need to be a little more careful. The key remark is that not only roots, but also their multiplicities, are preserved under linear fractional transformations (2.7). Indeed, substituting (2.17) in the transformation rule (2.8), we deduce that if p_0 is a root of multiplicity k and $\gamma p_0 + \delta \neq 0$, so that p_0 is not mapped to ∞, then $\bar{p}_0 = (\alpha p_0 + \beta)/(\gamma p_0 + \delta)$ will be a root of multiplicity k also. On the other hand, if $\gamma p_0 + \delta = 0$, then p_0 will map to an infinite root $\bar{p}_0 = \infty$ of the same multiplicity as p_0, in accordance with the following definition.

Definition 2.17. Let $Q(p)$ be an inhomogeneous binary form of degree n. The point $p_0 = \infty$ is said to be a *root* with *multiplicity* k if and only if the point $\bar{p}_0 = 0$ is a root of multiplicity k for the inverted polynomial $\overline{Q}(\bar{p}) = \bar{p}^{\,n} Q(1/\bar{p})$.

Exercise 2.18. Prove that an inhomogeneous binary form (2.2) has an infinite root of multiplicity k if and only if its leading k coefficients vanish: $a_n = a_{n-1} = \cdots = a_{n-k+1} = 0$. Thus the naïve degree of an inhomogeneous polynomial is strictly less than its degree if and only if the polynomial has an infinite root.

Inclusion of the roots at infinity completes the projective version of the Fundamental Theorem of Algebra.

Theorem 2.19. *An inhomogeneous polynomial $Q(p)$ of degree n has precisely n complex roots, counting multiplicities and roots at ∞. Moreover, the linear fractional transformation* (2.7) *maps each root of $Q(p)$ to a root having the same multiplicity of the transformed polynomial $\overline{Q}(\bar{p})$, as given by* (2.8).

For example, if $Q(p) = p^2 - 3p$ has degree 2, then it has two roots, namely 0 and 3. Under the inversion $p = 1/\bar{p}$, the transformed polynomial is $\overline{Q}(\bar{p}) = -3\bar{p} + 1$, which has corresponding roots $\infty = \frac{1}{0}$ and $\frac{1}{3}$.

Note that any polynomial can be readily transformed to one that does not have ∞ as a root and so is genuinely of degree n.

On the homogeneous level, the factorization of an inhomogeneous polynomial (2.17) of degree n translates into a complete factorization of its homogeneous counterpart into n linear factors. We will, for later convenience, always choose the points representing the homogenized roots so that the factorization takes the *normal form*

$$Q(x,y) = \prod_{\nu=1}^{n} (y_\nu x - x_\nu y)\,. \tag{2.18}$$

The normal factorization (2.18) requires that $\prod y_\nu = a_n$ be the leading coefficient of $Q(x,y)$, which can clearly be arranged by rescaling any one of the root representatives; leaving all the roots in general position gives a factorization of the same form (2.18), but with an additional nonzero coefficient in front of the product. If the horizontal line through $(1,0)$ is a root of $Q(x,y)$ having multiplicity k, so that the factor y^k appears in (2.18), then ∞ will be a root of multiplicity k of $Q(p)$, which must therefore satisfy the conditions of Exercise 2.18. Under the projective reduction (2.2) the "infinite" factors y^k all reduce to the constant 1, which accounts for the missing factors in (2.17).

Invariants and Covariants

We have now arrived at the key object of study in classical invariant theory — the concept of an invariant. Our motivational example is the discriminant (1.3) of a quadratic polynomial. The crucial property which we shall generalize is how it behaves under general linear (or, in the projective version, linear fractional) transformations. According to (1.13), the discriminant is not, strictly speaking, invariant, but rather is multiplied by a suitable power of the determinant of the matrix governing the transformation rule. This justifies the general definition.

Definition 2.20. An *invariant* of a binary form $Q(x,y)$ of degree n is a function $I(\mathbf{a}) = I(a_0,\dots,a_n)$, depending on its coefficients $\mathbf{a} = (a_0,\dots,a_n)$, which, up to a determinantal factor, does not change under the general linear transformation:

$$I(\mathbf{a}) = (\alpha\delta - \beta\gamma)^k\, I(\overline{\mathbf{a}}). \tag{2.19}$$

Here $\overline{\mathbf{a}} = (\bar a_0,\dots,\bar a_n)$ are the coefficients of the transformed polynomial (2.5), given explicitly in (2.6). The determinantal power $k = \operatorname{wt} I$ is called the *weight* of the invariant.

For example, according to (1.13), the discriminant of a binary quadratic is an invariant of weight 2. In fact, it is not difficult to prove that the discriminant is the only independent invariant of a quadratic polynomial, meaning that the other invariant is a power Δ^m of the discriminant, which has weight $k = 2m$.

While invariants are of fundamental importance in the geometry of binary forms, by themselves they do not paint the entire picture. Indeed, only when the discriminant of a quadratic form is nonzero does it completely determine its equivalence class and hence its canonical form. According to the table on p. 8, there are two possible canonical forms for quadratic polynomials with vanishing discriminant, either $Q(p) \equiv 1$ if the form is not identically zero and hence has a double root, or $Q(p) \equiv 0$. A similar situation holds for forms of higher degree, particularly for those with vanishing invariants, in which more subtle algebraic information is required than can be provided by the invariants. Classically, it was recognized that one needs to also consider functions depending not only on the coefficients of the binary form, but also on the independent variables x and y. This leads one to the more general definition of a "covariant".

Definition 2.21. A *covariant* of *weight* k of a binary form Q of degree n is a function $J(\mathbf{a}, \mathbf{x}) = J(a_0, \dots, a_n, x, y)$ depending both on the coefficients a_i and on the independent variables $\mathbf{x} = (x, y)$ which, up to a determinantal factor, is unchanged under linear transformations:

$$J(\mathbf{a}, \mathbf{x}) = (\alpha\delta - \beta\gamma)^k \, \bar{J}(\bar{\mathbf{a}}, \bar{\mathbf{x}}). \tag{2.20}$$

Note that invariants are just covariants that do not explicitly depend on $\mathbf{x}$. If the weight of a covariant (or invariant) J is $k = 0$, we call J an *absolute covariant*. The simplest covariant is the form Q itself, which in view of (2.5) forms an absolute covariant. For a binary quadratic, this is essentially the only covariant. More specifically, every polynomial covariant of a binary quadratic is given by a power product $J = \Delta^m Q^l$ depending on the form and its discriminant; the weight of J is $2m$. All of the important invariants and covariants are polynomial functions of the coefficients and the variables x, y, from which more general combinations, such as rational covariants, can be readily constructed.

The Simplest Examples

Let us now discuss some particular examples which will serve to illustrate and motivate the general features of the theory. Since we have already

exhausted the study of the binary quadratic, we now turn to cubic and quartic polynomials.

Example 2.22. Consider a binary cubic form

$$Q(\mathbf{x}) = a_3x^3 + 3a_2x^2y + 3a_1xy^2 + a_0y^3. \tag{2.21}$$

It turns out that there is just one fundamental invariant

$$\Delta = a_0^2a_3^2 - 6a_0a_1a_2a_3 + 4a_0a_2^3 - 3a_1^2a_2^2 + 4a_1^3a_3, \tag{2.22}$$

called the *discriminant* of the cubic Q. The direct proof that Δ is an invariant of weight 6 is a lengthy computation, but this will follow directly from the general theory presented below. The vanishing of Δ has an immediate geometric interpretation: $\Delta = 0$ if and only if Q has a double or a triple root; see Theorem 2.39.

The most important covariant of a cubic or, indeed, of any binary form is its *Hessian*

$$H = Q_{xx}Q_{yy} - Q_{xy}^2. \tag{2.23}$$

Here, and below, we shall often use subscripts to denote partial derivatives, so that

$$Q_x = \frac{\partial Q}{\partial x}, \qquad Q_y = \frac{\partial Q}{\partial y}, \qquad Q_{xx} = \frac{\partial^2 Q}{\partial x^2}, \qquad Q_{xy} = \frac{\partial^2 Q}{\partial x \partial y},$$

and so on. If Q has degree n, its Hessian will be a polynomial of degree $2n - 4$, whose coefficients depend quadratically on the coefficients of Q itself. Moreover, the Hessian forms a covariant of weight 2. The covariance of the Hessian will be shown in Chapter 5, although the interested reader might wish to try to prove this directly here. Note that the Hessian of a quadratic form is just 4 times its discriminant. If Q is a cubic, then its Hessian is a quadratic polynomial and is given explicitly by

$$\tfrac{1}{36}H = (a_1a_3 - a_2^2)\,x^2 + (a_0a_3 - a_1a_2)\,xy + (a_0a_2 - a_1^2)\,y^2. \tag{2.24}$$

The geometrical significance of the Hessian is contained in the following basic result; the general proof can be found on p. 91.

Proposition 2.23. *A binary form $Q(x, y)$ has vanishing Hessian, $H \equiv 0$, if and only if $Q(x, y) = (cx + dy)^n$ is the n^{th} power of a linear form.*

In particular, the Hessian of a cubic is identically zero if and only if the cubic has a triple root. (The reader is invited to prove this directly

from the explicit formula (2.24).) Since the quadratic Hessian (2.24) is a covariant, its discriminant, which is

$$\widetilde{\Delta} = \tfrac{1}{4}\,(a_0a_3 - a_1a_2)^2 - (a_1a_3 - a_2^2)(a_0a_2 - a_1^2), \tag{2.25}$$

is also an invariant for the cubic. (This is a special case of the general technique of composing covariants to be discussed shortly.) Expanding and comparing with (2.22), we see that this invariant is just a multiple of the original cubic discriminant: $\widetilde{\Delta} = \frac{1}{4}\Delta$. Consequently, a cubic has a multiple root if and only if its Hessian has a multiple root.

If K, L are any two covariants of a binary form Q, their *Jacobian*

$$J = \frac{\partial(K, L)}{\partial(x, y)} = K_xL_y - K_yL_x \tag{2.26}$$

is also a covariant. This result can be proved directly, but will again follow from more general considerations to be discussed in Chapter 5. If K has degree m and weight k, and L has degree l and weight j, then their Jacobian J has degree $m + l - 2$ and weight $k + j + 1$. For a binary cubic, it turns out that, besides the form Q and its Hessian, H, there is only one other independent covariant — the Jacobian of Q and H:

$$\begin{aligned} T &= Q_xH_y - Q_yH_x \\ &= -Q_yQ_{yy}Q_{xxx} + (2Q_yQ_{xy} + Q_xQ_{yy})Q_{xxy} - \\ &\qquad - (Q_yQ_{xx} + 2Q_xQ_{xy})Q_{xxy} + Q_xQ_{xx}Q_{yyy}. \end{aligned} \tag{2.27}$$

If Q is a binary cubic, then T is also a cubic polynomial whose coefficients have degree 3 in the coefficients of Q and forms a covariant of weight 3. A classical result, which we shall prove in Chapter 7, states that every polynomial invariant or covariant of a binary cubic can be written in terms of the covariants Q, H, T, and the invariant Δ.

Remark: Both the Hessian (2.23) and the Jacobian covariant (2.27) are homogeneous *differential polynomials* of the function Q, meaning that they can be expressed as polynomials in Q and its derivatives. In fact, *every* polynomial covariant and invariant of a binary form can be written as a homogeneous, constant coefficient differential polynomial. The First Fundamental Theorem 6.14 of classical invariant theory provides the explicit mechanism for constructing the differential polynomials that give rise to classical covariants and invariants.

Cubics can also be completely characterized by their covariants. Suppose first that the cubic has three distinct roots and so is characterized by the invariant condition $\Delta \neq 0$. In the complex case, we can

place them anywhere we like in $\mathbb{CP}^1$ by a suitable linear fractional transformation, e.g., $-1, 1$, and ∞, resulting in the canonical form $p^2 - 1$; see Example 4.30. The cubic has a double root if and only if its discriminant vanishes, but its Hessian is not identically zero; placing the double root at ∞ and the simple root at 0 leads to the canonical form p. A (nonzero) cubic has a single triple root if and only if its Hessian vanishes; the canonical form can be taken either to be p^3, by placing the root at 0, or to be 1, with the root sent to ∞. The complete list of complex canonical forms is given by the following table.

Canonical Forms for Complex Binary Cubics

I.	$p^2 - 1$	$\Delta \neq 0$	simple roots
II.	p	$\Delta = 0,\ H \not\equiv 0$	double root
III.	1	$H \equiv 0,\ Q \not\equiv 0$	triple root
IV.	0	$Q \equiv 0$	

In the real case, the first canonical form splits into two real forms, distinguished by the sign of its discriminant, depending on whether the cubic has any complex roots. If so, they can be placed at $\pm i$ and ∞ by a real linear fractional transformation. The remaining cases are unchanged. In this manner, we complete the classification of real canonical forms.

Canonical Forms for Real Binary Cubics

Ia.	$p^2 + 1$	$\Delta > 0$	two complex roots
Ib.	$p^2 - 1$	$\Delta < 0$	three simple real roots
II.	p	$\Delta = 0,\ H \not\equiv 0$	double root
III.	1	$H \equiv 0,\ Q \not\equiv 0$	triple root
IV.	0	$Q \equiv 0$	

Example 2.24. Consider next the binary quartic

$$Q(\mathbf{x}) = a_4x^4 + 4a_3x^3y + 6a_2x^2y^2 + 4a_1xy^3 + a_0y^4. \tag{2.28}$$

There are two fundamental invariants:

$$i = a_0 a_4 - 4a_1 a_3 + 3a_2^2, \tag{2.29}$$

which is of weight 4, and

$$j = \det \begin{vmatrix} a_4 & a_3 & a_2 \\ a_3 & a_2 & a_1 \\ a_2 & a_1 & a_0 \end{vmatrix}, \tag{2.30}$$

which is of weight 6. (Again, these remarks can be verified directly, but will follow more simply from the subsequent general theory.) The vanishing of the invariants i and/or j has geometric meaning: $i = j = 0$ if and only if Q has a triple or a quadruple root. Furthermore, if $i = 0$, $j \neq 0$, the roots form an "equi-anharmonic quadruplet", whereas $j = 0$ if and only if Q can be written as the sum of two fourth powers, $Q = (ap+b)^4 + (cp+d)^4$, and the roots form an "anharmonic quadruplet"; see Gurevich, [**97**; Exercise 25.7], for definitions and details. Note further that since i has weight 4 and j has weight 6, both i^3 and j^2 are relative invariants of weight 12. Therefore the ratio i^3/j^2 is an absolute invariant, and its value is fixed. Any linear combination of i^3 and j^2 is again a relative invariant of weight 12. The most important of these is the *discriminant* $\Delta = i^3 - 27j^2$, which vanishes if and only if the quartic has a multiple root; see below.

If Q is a quartic polynomial, then its Hessian (2.23) is also a quartic,

$$\begin{aligned} \tfrac{1}{144} H = {} & (a_2 a_4 - a_3^2)x^4 + 2(a_1 a_4 - a_2 a_3)x^3 y + \\ & + (a_0 a_4 + 2a_1 a_3 - 3a_2^2)x^2 y^2 + 2(a_0 a_3 - a_1 a_2)x y^3 + (a_0 a_2 - a_1^2)y^4, \end{aligned} \tag{2.31}$$

and is a covariant of weight 2. By Proposition 2.23, $H \equiv 0$ if and only if Q has a single quadruple root. As with the cubic, the only other

Canonical Forms for Complex Binary Quartics

I.	$p^4 + \mu p^2 + 1$	$\mu \neq \pm 2$, $\Delta \neq 0$	simple roots
II.	$p^2 + 1$	$\Delta = 0$, $T \not\equiv 0$	one double root
III.	p^2	$\Delta = 0$, $T \equiv 0$, $i \neq 0$	two double roots
IV.	p	$i = j = 0$, $H \not\equiv 0$	triple root
V.	1	$H \equiv 0$, $Q \neq 0$	quadruple root
VI.	0	$Q \equiv 0$	

independent covariant is the Jacobian $T = Q_x H_y - Q_y H_x$ of Q and H. As we shall prove in Chapter 7, every polynomial invariant or covariant of a binary quartic can be written in terms of the invariants i, j and the covariants Q, H, T. One can now use the invariants and covariants to provide a complete classification of binary quartics.

Exercise 2.25. Determine the real classification of binary quartics; see also [**97**; Exercises 25.13, 25.14].

Degree, Order, and Weight

Since the linear transformations (2.4) induce linear maps on the coefficients of a binary form, if J is any polynomial covariant, its homogeneous summands are individually polynomial covariants. Therefore we can, without loss of generality, restrict our attention to homogeneous covariants. We shall now make this requirement more precise and look at some elementary consequences.

Definition 2.26. Let $J(\mathbf{a}, \mathbf{x})$ be a homogeneous polynomial covariant of a binary form. The *degree* of J is its degree in the independent variables $\mathbf{x}$. The *order* of J is its degree in the coefficients $\mathbf{a}$ of the form.

So far we have been considering the case of a binary form that has weight zero, meaning that there is no extra determinantal factor in its transformation rules (2.5). More generally, we can assign a nonzero weighting to the original binary form.

Definition 2.27. A binary form $Q(\mathbf{x})$ is said to have *weight* m if, under the action of GL(2), its coefficients are subject to the transformation rules induced by the change of variables formula

$$\begin{aligned} Q(x, y) &= (\alpha\delta - \beta\gamma)^m \, \overline{Q}(\alpha x + \beta y, \gamma x + \delta y) \\ &= (\alpha\delta - \beta\gamma)^m \, \overline{Q}(\bar{x}, \bar{y}). \end{aligned} \tag{2.32}$$

Since reweighting a binary form only introduces an additional determinantal factor, all the homogeneous invariants and covariants of a weight 0 binary form remain invariants and covariants of a weight m form, albeit with a suitably modified weighting.

Proposition 2.28. *If $J(\mathbf{a}, \mathbf{x})$ is a homogeneous covariant of weight k and order j for a binary form Q of weight 0, then J will be a covariant of weight $k + jm$ and order j when Q has weight m.*

In particular, the Hessian $H = Q_{xx}Q_{yy} - Q_{xy}^2$ of a weight m form will have the modified weight $2 + 2m$. An interesting example, which turns out to be important for the study of differential operators, cf. [**169, 235**], is the case of a quartic polynomial of weight -2, so that its transformation rule includes the reciprocal of the square of the determinant. In this case, the invariants i and j both have weight 0, i.e., they are absolute invariants.

The degree, order, and weight of a covariant are intimately related, which implies that any two of these uniquely determine the third.

Proposition 2.29. *Let $J(\mathbf{a}, \mathbf{x})$ be a homogeneous polynomial covariant of a binary form $Q(\mathbf{x})$. Then*

$$\deg J + 2 \operatorname{wt} J = (\deg Q + 2 \operatorname{wt} Q) \operatorname{ord} J. \tag{2.33}$$

Proof: Let $n = \deg Q$, $m = \operatorname{wt} Q$, $j = \operatorname{ord} J$, $k = \operatorname{wt} J$, $i = \deg J$. Then, by homogeneity, $J(\mu\mathbf{a}, \nu\mathbf{x}) = \mu^j \nu^i J(\mathbf{a}, \mathbf{x})$. On the other hand, consider the effect of a scaling transformation $\overline{\mathbf{x}} = \lambda\mathbf{x}$, which has determinant $\det A = \lambda^2$. According to (2.32), the coefficients of the transformed polynomial $\overline{Q}$ are given by $\overline{\mathbf{a}} = \lambda^{-n-2m}\mathbf{a}$. The covariance of J implies that

$$J(\mathbf{a}, \mathbf{x}) = \lambda^{2k} J(\overline{\mathbf{a}}, \overline{\mathbf{x}}) = \lambda^{2k} J(\lambda^{-n-2m}\mathbf{a}, \lambda\mathbf{x}) = \lambda^{i+2k-(n+2m)j} J(\mathbf{a}, \mathbf{x}).$$

Consequently, the final exponent of λ in this equation must vanish, which suffices to prove (2.33). *Q.E.D.*

Exercise 2.30. Prove that a binary form of even degree has no nonzero polynomial covariants of odd degree. Prove that every nonzero polynomial covariant of a binary form of odd degree is either of even order and even degree or of odd order and odd degree.

Exercise 2.31. Let Q be a binary form of degree n and weight 0, with coefficients $\mathbf{a} = (a_0, \dots, a_n)$. Suppose $I = \sum c_M \mathbf{a}^M$ is an invariant of weight k and order j with constituent monomials $\mathbf{a}^M = (a_0)^{m_0}(a_1)^{m_1} \cdots (a_n)^{m_n}$. Prove that

$$\begin{aligned} j &= m_0 + m_1 + m_2 + \cdots + m_n, \\ k = \tfrac{1}{2}nj &= m_1 + 2m_2 + 3m_3 + \cdots + nm_n \\ &= nm_0 + (n-1)m_1 + (n-2)m_2 + \cdots + m_{n-1}. \end{aligned} \tag{2.34}$$

Next, write $n = 2l$ or $n = 2l + 1$ depending on whether Q is of even or odd degree. Divide the coefficients into two subsets $\mathbf{a}_- = (a_0, \dots, a_l)$ and $\mathbf{a}_+ = (a_{n-l}, \dots, a_n)$. (Note that a_l appears in both subsets when

$n = 2l$ is even.) Prove that every term a^M in the invariant I must contain at least one factor from $\mathbf{a}_-$ and at least one factor from $\mathbf{a}_+$. In other words, no term in an invariant can depend solely on the coefficients $\mathbf{a}_-$ or solely on the coefficients $\mathbf{a}_+$. Is a similar result true for covariants?

Construction of Covariants

A wide variety of useful techniques for constructing covariants of binary forms have been proposed, including algebraic methods, symbolic methods, methods using differential polynomials and/or differential invariants, infinitesimal methods, methods based on the roots of the polynomials, and representation-theoretic methods. We begin by looking at the simplest algebraic methods that can be used to construct covariants.

The most trivial method is to multiply covariants. If J is a covariant of weight j and K has weight k, then the product $J \cdot K$ is a covariant of weight $j + k$. Therefore we can take general products of (powers of) covariants to straightforwardly construct other covariants, trivially related to the original covariants. However, these are typically not of great interest as they provide essentially the same information as their constituents. It is also possible to add covariants, but *only if they have the same weight*. Thus, $J+K$ will be a covariant if and only if both J and K have equal weight j, in which case their sum (or any other constant coefficient linear combination thereof) also has weight j. For example, if we begin with a binary quartic, having the standard (classical) weight 0, then its invariants i and j, cf. (2.29), (2.30), have respective weights 4 and 6, so $i+j$ is *not* an invariant since its components are multiplied by different determinantal powers. (However, its value is invariant if we only allow unimodular linear transformations.) The powers i^3 and j^2 have weight 12, and so any linear combination, including the discriminant $\Delta = i^3 - 27j^2$, is also an invariant of weight 12. On the other hand, if we give the original quartic weight -2, then, as remarked earlier, both i and j have weight 0, and so the sum $i + j$ is also an absolute invariant for this special weighting; indeed, so is any function $F(i, j)$.

A second method for constructing covariants, alluded to in our discussion of the binary cubic, is the method of composition. If $Q(\mathbf{x})$ is a binary form with coefficients $\mathbf{a}$, any polynomial covariant $J(\mathbf{a}, \mathbf{x})$ can itself be considered as a binary form, whose weight is the weight of J. Let $\mathbf{b} = \varphi(\mathbf{a})$ denote the coefficients of J, which are certain polynomials in the coefficients $\mathbf{a}$ of Q. It is not hard to see that if $K(\mathbf{b}, \mathbf{x})$

is a covariant depending on the coefficients of J, then the polynomial $\widetilde{K}(\mathbf{a}, \mathbf{x}) = K(\varphi(\mathbf{a}), \mathbf{x})$ obtained by replacing the coefficients of J by their formulae in terms of the coefficients of Q provides a covariant of the original form.

For example, if Q is a quartic polynomial (2.28), then its Hessian $H = H(Q)$ is itself a quartic polynomial, cf. (2.23). Thus, the i and j invariants of H, denoted $i \circ H = i(H(Q))$ and $j \circ H = j(H(Q))$, will in turn yield new invariants of Q. To compute these, we replace the coefficient a_i of $x^i y^{4-i}$ in (2.29) and (2.30) by the corresponding coefficients b_i of $x^i y^{4-i}$ in H itself, so, for instance, a_0 is replaced by $144(a_0 a_2 - a_1^2)$, and so on. However, if we already know that i and j are the only independent invariants of Q, it will not be surprising that we discover that these new invariants can be re-expressed in terms of i and j. For instance, $i \circ H = 1728\, i^2$. In Chapter 6 we shall discover more efficient methods for determining such identities.

Exercise 2.32. Determine the general rule for the behavior of weights under composition of covariants.

Joint Covariants and Polarization

More generally, if we are given a system $Q_1(\mathbf{x}), \dots, Q_l(\mathbf{x})$ of homogeneous polynomials, their common or correlated geometrical properties will be classified by their *joint invariants* and *covariants*. By definition, these are functions $J(\mathbf{a}^1, \dots, \mathbf{a}^l, \mathbf{x})$ depending on all the coefficients $\mathbf{a}^\kappa = (\dots a_i^\kappa \dots)$ of the Q_κ, and, in the case of covariants, the variables $\mathbf{x} = (x, y)$, which, when all the forms are simultaneously subjected to a linear transformation, satisfy the same basic transformation rule (2.20). The determinantal power k is, as before, the weight of the joint covariant. The forms themselves may be of varying weights, the most common case occurring when they all have weight 0. We shall say that the joint covariant has *order* $i = (i_1, \dots, i_l)$ if it is a homogeneous function of degree i_κ in the coefficients $\mathbf{a}^\kappa$ of Q_κ. The most important joint covariants typically arise as differential polynomials $J = J[Q_1, \dots, Q_l]$ depending on the forms and their derivatives.

For example, if $Q(x, y) = ax + by$, $R(x, y) = cx + dy$ are linear forms of weight 0, their determinant $ad - bc$ is a bilinear, i.e., order $(1, 1)$, joint invariant of weight 1. This is a special case of the general Jacobian covariant $J = Q_x R_y - Q_y R_x$ already considered in (2.26). Another

important example is the bilinear (or polarized) version of the Hessian,

$$H[Q,R] = Q_{xx}R_{yy} - 2Q_{xy}R_{xy} + Q_{yy}R_{xx}; \tag{2.35}$$

the Hessian itself, (2.23), is recovered by setting $Q = R$. If Q and R have weight 0, then $H[Q,R]$ has weight 2. As in the case of a single form, if $J = J[Q_1, \dots, Q_l]$ is joint covariant of order $(i_1, \dots, i_l)$ and weight k when each Q_α has weight 0, then J remains a joint covariant of weight $k + \sum i_\kappa m_\kappa$ when Q_κ has revised weight m_κ.

The connection between the Hessian and its polarized counterpart is a special case of a general procedure for relating joint covariants and ordinary covariants, first noted in Boole's original paper, [**24**]. In the simplest version, suppose $K(\mathbf{a}, \mathbf{x})$ is any polynomial depending on the coefficients $\mathbf{a} = (a_0, \dots, a_n)$ of the degree n binary form Q. Define its *polarization* to be the joint polynomial

$$J(\mathbf{a}, \mathbf{b}, \mathbf{x}) = \sum_{i=0}^{n} b_i \, \frac{\partial K}{\partial a_i}(\mathbf{a}, \mathbf{x}), \tag{2.36}$$

depending on the respective coefficients $\mathbf{a}$, $\mathbf{b}$ of two binary forms Q, R *of the same degree*. If $K = K[Q]$ is a differential polynomial in Q, then its polarization $J[Q,R]$ is obtained by formally applying the differentiation process $R\,\partial/\partial Q$ to J. The formal differential operator $\partial/\partial Q$ does not affect the x, y coordinates, or derivatives with respect to them. For example, if $K = QQ_yQ_{xxy}$, then

$$J = R\frac{\partial K}{\partial Q} = RQ_yQ_{xxy} + QR_yQ_{xxy} + QQ_yR_{xxy},$$

while the polarization of the Hessian (2.23) is precisely (2.35).

Given a joint function $J[Q,R]$ depending on two binary forms of the *same* degree, we define its *trace* to be the function $K[Q] = J[Q,Q]$ obtained by setting $Q = R$. If $J[Q,R]$ is a joint covariant, then its trace is an ordinary covariant. The trace operation is, in a sense, the inverse process to polarization. If $K[Q]$ has order k, and $J[Q,R] = R\,\partial K/\partial Q$ is its polarization, then Euler's formula (2.13) implies that the trace of J recovers the original function up to a multiple: $J[Q,Q] = k\,K[Q]$. For example, setting $Q = R$ in (2.35) gives twice the Hessian covariant. On the other hand, the trace of the Jacobian joint covariant is trivial, and so one cannot obtain it by polarizing an ordinary covariant.

Proposition 2.33. *If $K[Q]$ is a covariant of weight k and order l for the single binary form Q of degree n and weight m, then its polarization $J[Q,R] = R\,\partial K/\partial Q$ is a joint covariant of weight k and order*

$(l-1,1)$ for the pair of degree n, weight m forms Q, R. Conversely, if $J[Q,R]$ is a joint covariant of weight k and order (i,j) for two forms Q, R, of the same degree and weight, then its trace $K[Q] = J[Q,Q]$ is a weight k and order $i+j$ covariant for the single form Q.

Both polarization and trace can be readily generalized to joint covariants $J[Q_1, \dots, Q_l]$ depending on several forms. If the Q_κ's all have the same degree and weight, then the trace $J[Q, \dots, Q]$, which is obtained by equating all the forms, is a covariant of the single form Q of the given degree and weight. One can also take partial traces by equating only some of the forms. Conversely, if Q_α and Q_β have the same degree, then the general *polarization process*

$$\widetilde{J}[Q_1, \dots, Q_l] = Q_\beta \frac{\partial J}{\partial Q_\alpha}[Q_1, \dots, Q_l] \tag{2.37}$$

defines another covariant, whose order in Q_β has increased by one, and whose order in Q_α has decreased by one. One can iterate this procedure to provide joint covariants depending on more and more forms (all of the same degree). An important problem then is to find a minimal system of joint covariants, from which all others can be constructed by polarization and algebraic operations. See Weyl, [**231**; p. 251], and Chapter 8 for further results in this direction.

Exercise 2.34. Find the general formula, analogous to (2.33), for the weight of a joint covariant.

Resultants and Discriminants

A particularly important joint invariant of two polynomials is their resultant, which indicates the existence of common roots. Let

$$\begin{aligned} P(\mathbf{x}) &= \widetilde{a}_m x^m + \widetilde{a}_{m-1} x^{m-1} y + \cdots + \widetilde{a}_0 y^m, \\ Q(\mathbf{x}) &= \widetilde{b}_n x^n + \widetilde{b}_{n-1} x^{n-1} y + \cdots + \widetilde{b}_0 y^n, \end{aligned} \tag{2.38}$$

be homogeneous polynomials of respective degrees m and n. (The formulae are a bit easier to read if we omit our usual binomial coefficients.) If P and Q have a common nonconstant factor F, then we can write $P = F \cdot R$, $Q = F \cdot S$, and hence

$$S(\mathbf{x})\, P(\mathbf{x}) = R(\mathbf{x})\, Q(\mathbf{x}), \quad \text{where} \quad \begin{matrix} \deg R < \deg P, \\ \deg S < \deg Q, \end{matrix} \quad R, S \not\equiv 0. \tag{2.39}$$

The unique prime factorization of complex polynomials implies that (2.39) is necessary and sufficient for the existence of a common factor. In fact, by multiplying both sides by a common nonzero factor, we can assume that the degrees of R and S are precisely one less than those of P and Q, and so

$$\begin{aligned} R(\mathbf{x}) &= \widetilde{c}_{m-1}x^{m-1} + \widetilde{c}_{m-2}x^{m-2}y + \cdots + \widetilde{c}_0 y^{m-1}, \\ S(\mathbf{x}) &= \widetilde{d}_{n-1}x^{n-1} + \widetilde{d}_{n-2}x^{n-2}y + \cdots + \widetilde{d}_0 y^{n-1}. \end{aligned} \tag{2.40}$$

Substituting into (2.39) and equating the various coefficients of powers of x, y, we deduce that the coefficients of R, S must satisfy the system of $m + n$ linear equations

$$\begin{aligned} \widetilde{a}_m\widetilde{d}_{n-1} &= \widetilde{b}_n\widetilde{c}_{m-1}, \\ \widetilde{a}_m\widetilde{d}_{n-2} + \widetilde{a}_{m-1}\widetilde{d}_{n-1} &= \widetilde{b}_n\widetilde{c}_{m-2} + \widetilde{b}_{n-1}\widetilde{c}_{m-1}, \\ \widetilde{a}_m\widetilde{d}_{n-3} + \widetilde{a}_{m-1}\widetilde{d}_{n-2} + \widetilde{a}_{m-2}\widetilde{d}_{n-1} &= \widetilde{b}_n\widetilde{c}_{m-3} + \widetilde{b}_{n-1}\widetilde{c}_{m-2} + \widetilde{b}_{n-2}\widetilde{c}_{m-1}, \\ &\vdots \\ \widetilde{a}_0\widetilde{d}_0 &= \widetilde{b}_0\widetilde{c}_0. \end{aligned} \tag{2.41}$$

If this homogeneous system of linear equations is to have a nonzero solution, then its associated $(m + n) \times (m + n)$ coefficient matrix must have zero determinant. The determinant† in question is

$$\mathbf{R}[P, Q] = \det \begin{vmatrix} \widetilde{a}_m & \widetilde{a}_{m-1} & \cdots & \cdots & \widetilde{a}_0 & & & \\ & \widetilde{a}_m & \widetilde{a}_{m-1} & \cdots & \cdots & \widetilde{a}_0 & & \\ & & \ddots & & & & \ddots & \\ & & & \widetilde{a}_m & \widetilde{a}_{m-1} & \cdots & \cdots & \widetilde{a}_0 \\ \widetilde{b}_n & \widetilde{b}_{n-1} & \cdots & \cdots & \widetilde{b}_0 & & & \\ & \widetilde{b}_n & \widetilde{b}_{n-1} & \cdots & \cdots & \widetilde{b}_0 & & \\ & & \ddots & & & & \ddots & \\ & & & \widetilde{b}_n & \widetilde{b}_{n-1} & \cdots & \cdots & \widetilde{b}_0 \end{vmatrix}, \tag{2.42}$$

in which there are n rows of $\widetilde{a}$'s and m rows of $\widetilde{b}$'s, and all blank spaces are 0. The resulting polynomial in the coefficients of P and Q is called their *resultant*. For example, the resultant of a quadratic $P = ax^2 + 2bxy + cy^2$

† We have, in accordance with standard practice, [**222**; §5.8], transposed the coefficient matrix of (2.41).

and a linear form $Q = dx + ey$ is the 3×3 determinant

$$\mathbf{R} = \det \begin{pmatrix} a & 2b & c \\ d & e & 0 \\ 0 & d & e \end{pmatrix} = ae^2 - 2bde + cd^2,$$

which vanishes if and only if $[-e, d]$ is a homogenized root of P.

Theorem 2.35. *The resultant of two polynomials vanishes if and only if they have a common nonconstant factor and hence have a common complex (possibly infinite) root.*

There is an alternative formula for the resultant in terms of the roots of the polynomials, which immediately proves its invariance under linear transformations.

Theorem 2.36. *Let P have homogenized roots $\mathbf{x}_1, \dots, \mathbf{x}_m$ and Q homogenized roots $\widehat{\mathbf{x}}_1, \dots, \widehat{\mathbf{x}}_n$, both of which are taken in normal factored form* (2.18). *Then the resultant of P and Q can be written as the product of the differences of the roots*

$$\mathbf{R}[P, Q] = \prod_{\alpha=1}^{m} \prod_{\beta=1}^{n} (x_\alpha \widehat{y}_\beta - y_\alpha \widehat{x}_\beta) = \prod_{\alpha=1}^{m} Q(\mathbf{x}_\alpha) = (-1)^{mn} \prod_{\beta=1}^{n} P(\widehat{\mathbf{x}}_\beta). \tag{2.43}$$

Proof: Let us regard $R = \mathbf{R}[P, Q]$ as a polynomial function of the roots $\mathbf{x}_\alpha$, $\widehat{\mathbf{x}}_\beta$ of P and Q. Since $R = 0$ whenever two roots coincide, $\mathbf{x}_\alpha = \widehat{\mathbf{x}}_\beta$, it must admit the linear polynomial $x_\alpha \widehat{y}_\beta - y_\alpha \widehat{x}_\beta$ as a factor. The degree of R in the roots equals the degree of the product of all these factors, and hence R is a constant multiple of the right-hand side in (2.43). Our assumption that P and Q are in normal factored form can be used to show that the constant must be 1. *Q.E.D.*

Corollary 2.37. *If P, Q have respective degrees m, n and weights j, k, then the resultant $\mathbf{R}[P, Q]$ is a joint invariant of weight $mn+mk+nj$.*

Exercise 2.38. The k^{th} *subresultant* $R_k = \mathbf{R}_k[P, Q]$ of the polynomials P, Q is the $(m + n - 2k) \times (m + n - 2k)$ determinant obtained by deleting the first and last k rows and columns from the resultant determinant (2.42). Prove that P and Q have precisely k roots in common (counting multiplicities) if and only if their first k subresultants $R_0 = R, R_1, \dots, R_{k-1}$ vanish, while $R_k \neq 0$; see also [**23**; p. 197].

The *discriminant* of a binary form $Q(\mathbf{x})$ of degree n is, up to a

factor, just the resultant of Q and its derivative, namely,

$$\Delta[Q] = \frac{\mathbf{R}[Q, Q_x]}{n^n\, a_n} = \frac{\mathbf{R}[Q, Q_y]}{n^n\, a_0}. \tag{2.44}$$

Note that we can identify $Q_x = \partial Q/\partial x$ with the derivative $Q'(p)$ of the inhomogeneous version $Q(p) = Q(p, 1)$. Theorem 2.35 implies that the discriminant will detect the presence of common roots of $Q(p)$ and $Q'(p)$. These are precisely the multiple roots of Q.

Theorem 2.39. *Let Q be written in the normal factored form* (2.18) *with roots $\mathbf{x}_1, \dots, \mathbf{x}_n$. The discriminant of Q equals the product of the squares of the differences of the roots*

$$\Delta[Q] = \prod_{\alpha \neq \beta} (x_\alpha y_\beta - y_\alpha x_\beta)^2. \tag{2.45}$$

The discriminant vanishes if and only if Q has a multiple root. Moreover, if Q has degree n and weight m, then its discriminant is an invariant of weight $(n-1)(n+2m)$.

Proof: We compute the derivative of the factored form directly:

$$Q(x, y) = \prod_{\alpha=1}^{n} (x y_\alpha - y x_\alpha), \qquad \text{so} \qquad \frac{\partial Q}{\partial x} = \sum_{\beta=1}^{n} \prod_{\alpha \neq \beta} (x y_\alpha - y x_\alpha).$$

Substituting into the final expression in (2.43) produces (2.45). *Q.E.D.*

For example, the discriminant of a binary quadratic (1.7) is

$$\Delta = \frac{1}{4a_2} \det \begin{vmatrix} a_2 & 2a_1 & a_0 \\ 2a_2 & 2a_1 & 0 \\ 0 & 2a_2 & 2a_1 \end{vmatrix} = a_0 a_2 - a_1^2,$$

which is $\frac{1}{4}$ times its Hessian. Similarly, the discriminant of a binary cubic is

$$\Delta = \frac{1}{27a_3} \det \begin{vmatrix} a_3 & 3a_2 & 3a_1 & a_0 & 0 \\ 0 & a_3 & 3a_2 & 3a_1 & a_0 \\ 3a_3 & 6a_2 & 3a_1 & 0 & 0 \\ 0 & 3a_3 & 6a_2 & 3a_1 & 0 \\ 0 & 0 & 3a_3 & 6a_2 & 3a_1 \end{vmatrix}.$$

Expanding the determinant, we find that this agrees with the previous formula (2.22).

Exercise 2.40. Prove that the discriminant of a quartic is equal to the particular combination $i^3 - 27j^2$ of the invariants (2.29), (2.30).

The Hilbert Basis Theorem

Since appropriately homogeneous polynomial combinations of covariants are also covariants, an important algebraic problem is to find a *minimal* list of "fundamental" polynomial covariants, known as a Hilbert basis, that generate all others. Knowledge of a Hilbert basis for a given system of forms allows one to straightforwardly describe all covariants, and hence (presumably) all the intrinsic geometric properties of such forms.

Definition 2.41. Suppose $Q_1, \dots, Q_l$ are a collection of binary forms. A finite collection of invariants $I_1, \dots, I_m$ forms a *Hilbert basis* if every other invariant can be written as a polynomial function of the basis invariants: $I = P(I_1, \dots, I_m)$. Similarly, a finite collection of covariants $J_1, \dots, J_k$ forms a *Hilbert basis* if every other covariant J can be written as a polynomial in the basis covariants: $J = P(J_1, \dots, J_k)$.

For example, a Hilbert basis for the covariants of a binary quadratic consists of the form Q itself, and its discriminant, which is the only independent invariant. A binary cubic has 4 fundamental covariants, consisting of Q, the Hessian H, the Jacobian covariant T given in (2.27), and the discriminant Δ, which is the one invariant. A quartic has two invariants, i, j, and three covariants Q, H, T. These results, as well as those for the quintic and sextic, were known to Cayley, who then stated, [**41**], that binary forms of degree 7 or more do not have a finite Hilbert basis for their invariants. In 1868, Gordan, [**86**], succeeded in proving his finiteness theorem, which meant that Cayley was mistaken — *every* binary form admits a Hilbert basis. Gordan's method of proof is constructive, and so, at least in principle, one was now able to produce complete systems of invariants and covariants for general binary forms. However, Gordan's method has only been completely carried out for binary forms of degree at most 8, cf. [**92**; p. 132], [**226**]. The number of polynomial independent covariants rapidly increases with the order of the form, and the implementation of the method in higher degrees becomes infeasible (although modern computer algebra packages might come to the rescue). For example, quintic forms have 4 invariants and 23 covariants (including the invariants) in a complete Hilbert basis; while sextics have 5 invariants and 26 covariants.

Sylvester, [**208, 209**], produced tables of Hilbert bases for the covariants of binary forms of degree $n \leq 10$ and $n = 12$. However, as shown by Dixmier and Lazard, [**60, 61**], Sylvester's entry for the form of degree 7 is not correct — he misses several invariants, and so the

higher order computations are rather suspect. (Indeed, Sylvester makes several remarks about the anomalous nature of the binary septic — the case that also led Cayley astray — but does not conclude that his calculations are incorrect.) Sylvester does get the invariants correct for a form of degree 8, as was re-proved by Shioda, [**194**], but I do not know whether Sylvester's list of covariants is correct. For historical interest, Sylvester's tables, with known corrections in parentheses, are as follows. The proliferation of invariants and covariants at the higher orders is striking. However, one should trust the listed number of covariants only up to degree 6 and invariants only up to degree 8.

degree	2	3	4	5	6	7	8	9	10	12
# invariants	1	1	2	4	5	26 (30)	9	89	104	109
# covariants	2	4	5	23	26	124 (130)	69	415	475	949

Following Gordan's triumph with binary forms, the focus shifted to polynomials in three or more homogeneous variables. (See Chapter 10 for the precise definitions of invariants and covariants in the multivariate context.) Progress was slow, until the mathematical world was stunned when David Hilbert, at age 26, [**105**], suddenly and unexpectedly proved the existence of a Hilbert basis for *any* number of forms in *any* number of variables. Hilbert's celebrated theorem is the following:

Theorem 2.42. *Any finite system of homogeneous polynomials admits a Hilbert basis for its invariants, as well as for its covariants.*

Hilbert's original proof of the Finiteness Theorem was existential, thereby provoking Gordan's famous (perhaps apocryphal) exclamation "Das ist Theologie und nicht Mathematik." In response to such criticisms, Hilbert published a second, more difficult constructive proof, [**106**], although this is less well known, and Hilbert has been unjustly saddled with the reputation of killing off constructive invariant theory.† As recently emphasized by Sturmfels, [**204**], Hilbert's second proof, combined with the modern theory of Gröbner bases, [**28**, **54**], has the potential to be formed into a constructive algorithm for producing the Hilbert

† This and other invariant-theoretic apocrypha can mostly be traced to Weyl's incomplete and at times misleading historical remarks, [**231**; p. 27].

basis of a general system of forms. However, the actual implementation has yet to be completed, and there are counterclaims, [**226**], that one can "effectively" complete the classification only in the known cases.

Exercise 2.43. Let Q be a binary form. According to the general composition method, if J, K are any two covariants, their resultant $\mathbf{R}[J, K]$ and discriminants $\Delta[J]$ and $\Delta[K]$ will be invariants of Q. Consider the particular case when Q is a binary cubic. Since its discriminant $\Delta = \Delta[Q]$ is the only independent invariant, all such composed resultants and discriminants must be constant multiples of suitable powers of Δ. Prove the following formulae:

$$\begin{aligned} \Delta[Q] &= \Delta, & \Delta[H] &= -324\,\Delta, & \Delta[T] &= 2^8\,3^{12}\,\Delta^3, \\ \mathbf{R}[Q,H] &= 6^6\,\Delta^2, & \mathbf{R}[Q,T] &= -6^9\,\Delta^3, & \mathbf{R}[H,T] &= 2^{10}3^{12}\,\Delta^3. \end{aligned} \tag{2.46}$$

(A computer algebra package might come in handy.) Discuss implications for the possible joint root configurations of the cubic and its covariants.

Syzygies

While polynomial independence of the fundamental covariants appearing in a Hilbert basis has received the lion's share of interest in the algebraic approach to invariant theory, applications to geometry do not typically require such detailed, elusive information. Indeed, if one relaxes the requirement of polynomial independence to either rational, algebraic, or, most generally, functional independence, then complete results are much easier to obtain.

Although the number of independent invariants in a Hilbert basis of a binary form increases rapidly with its degree, a simple dimension count based on the orbits shows that the number of functionally independent invariants (for a generic form) cannot exceed $n - 2$, where $n \geq 3$ is the degree of the form; see Chapter 9. In Chapter 7, we will construct an explicit rational basis consisting of n rationally independent covariants, having the property that any other invariant or covariant can be written as a rational function thereof. Even better, in Chapter 8, we will find that a complete solution to the equivalence and symmetry problems for binary forms can be based on merely two absolute rational covariants, which involve only three particular polynomial covariants. This has the remarkable implication that the complete geometry of any binary form is encapsulated in these two covariants and their functional dependencies!

The reason that one does not require so many rationally or functionally independent covariants is that there exist certain polynomial identities, known as *syzygies*, among the basis covariants. For example, the four covariants of a binary cubic are related by the single syzygy

$$T^2 = 2^4 3^6 \Delta Q^2 - H^3. \tag{2.47}$$

Therefore, if one is willing to forgo reliance on polynomial covariants, one needs to understand only three of the cubic covariants. Similarly, the covariants of a quartic are also related by a single syzygy

$$T^2 = -\tfrac{16}{9} H^3 + 2^{10} 3^2 i Q^2 H - 2^{14} 3^4 j Q^3. \tag{2.48}$$

Methods for deriving such identities will be discussed in detail later.

One important application of the syzygy (2.47) is the following rather pretty solution to a general cubic equation $Q(p) = 0$. Suppose first that the discriminant $\Delta \neq 0$, so the cubic has three simple roots. We factor the syzygy as

$$H^3 = (108\sqrt{\Delta}\, Q - T)\,(108\sqrt{\Delta}\, Q + T). \tag{2.49}$$

The two cubic factors $108\sqrt{\Delta}\, Q \pm T$ do not have a common linear factor, since if they did, then T and Q would have a common root, and hence their resultant $\mathbf{R}[Q, T]$ would vanish. But, according to (2.46), the resultant is a multiple of Δ^3, and we assumed that $\Delta \neq 0$. Moreover, again by (2.46), the discriminant of the Hessian does not vanish, and so we factor the quadratic Hessian into *distinct* linear factors: $H = L \cdot M$. Equation (2.49) implies that (perhaps by relabeling the linear factors)

$$108\sqrt{\Delta}\, Q - T = L^3, \qquad 108\sqrt{\Delta}\, Q + T = M^3,$$

and hence Q is expressed as a sum of two cubes. This expression can be directly factored:

$$Q = \frac{L^3 + M^3}{216\sqrt{\Delta}} = \frac{(L + M)(L + \varepsilon M)(L + \varepsilon^2 M)}{216\sqrt{\Delta}}, \tag{2.50}$$

where $\varepsilon = \sqrt[3]{1}$ is a primitive cube root of unity.

On the other hand, if $\Delta = 0$, but $H \not\equiv 0$, then (2.46) implies that the discriminant of the quadratic Hessian vanishes, and so $H = L^2$ is a perfect square. Moreover, since the resultant of Q and H vanishes, Q admits L as a factor; in fact, it is not hard to see that the linear form L provides the double root of $Q = L^2 \cdot M$, and hence the solution is straightforward. The final nonzero case is when $H \equiv 0$, in which case $Q = L^3$ is a perfect cube, and immediately solvable.

Exercise 2.44. Use the quartic syzygy (2.48) to solve the general quartic equation. *Hint*: You will need to apply the preceding solution to a cubic equation; see [**107**; Lecture XXII].

A modern result, due to Hochster and Roberts, [**111**], states that the ring of covariants of a (system of) binary forms has the structure of a *Cohen–Macaulay domain*. A precise definition of this more subtle algebraic concept would be out of place here; the interested reader can consult [**204, 205**] for details. An important consequence of this result is the existence of a *Hironaka decomposition* of the ring of covariants; this means that every covariant can be written, uniquely, in the form

$$I = \Phi_0(I_1, \dots, I_n) + \sum_{\nu=1}^{k} \Phi_\nu(I_1, \dots, I_n)\, J_\nu, \tag{2.51}$$

where $I_1, \dots, I_n$ are algebraically independent covariants, and $J_1, \dots, J_k$ are additional covariants needed to complete the Hilbert basis. In fact, for a single binary form of degree n, the number of algebraically independent covariants equals the degree of the form. However, the number of auxiliary covariants is not known except in low order cases. For example, in the case of a cubic, in view of the syzygy (2.47), any covariant can be written as $C = \Phi(Q, H, \Delta) + \Psi(Q, H, \Delta)\, T$, while for a quartic, $C = \Phi(Q, H, i, j) + \Psi(Q, H, i, j)\, T$. A similar result holds for the rings of invariants, although this only becomes nontrivial for forms of degree 5 or more.

Chapter 3

Groups and Transformations

Following our preliminary foray into the basic ideas of invariant theory, it is now time to understand, in more detail, the mathematical foundations of our subject. Of course, one could continue to focus solely on invariant theory, but the full ramifications of our investigations would remain obscure without a proper appreciation for the underlying, modern mathematical theories, most of which can trace their genesis back to the problems of classical invariant theory itself. This chapter is devoted to a brief survey of the basic theory of transformation groups, starting with the properties of groups themselves. For our purposes, the most important examples are provided by simple actions on a linear space and their projective counterparts. Although our primary focus is on certain infinite, continuous groups, the present chapter will develop the general theory, which includes finite, discrete, infinite, and topological groups. More advanced methods that rely on the additional analytic structure of Lie groups will be postponed until Chapters 8 and 9.

Basic Group Theory

The theory of groups has its origins in the classical work of Lagrange, Abel, and Galois on the solubility of polynomials. (See, for example, [**237**, **239**], for historical surveys of group theory.) These mathematical giants discovered that the symmetries of a geometric object (in their case, the object was the set of roots to a polynomial equation) admit a certain underlying structure, which is crystallized in the definition of a "group". A half century later, Felix Klein clarified the foundational role of groups in geometry, and his justly famous *Erlanger Programm*, [**128**], showed how each type of geometry (Euclidean geometry, affine geometry, projective geometry, etc.) is completely characterized by an underlying transformation group. Simultaneously, motivated by the study of partial differential equations, Sophus Lie introduced and developed the theory of continuous or Lie groups, [**138**], which are manifested as symmetry groups of the solutions to a differential equation. Groups are ubiquitous

in mathematics and have an astounding variety of applications — to physics, to mechanics, to computer vision, to biology, and elsewhere. As quoted in [**239**], Alexandroff proclaims that "... the concepts of *number*, *set*, *function* and *group* are the four cornerstones on which the entire edifice of modern mathematics rests and to which any other mathematical concept reduces".

We begin our presentation with the fundamental definition.

Definition 3.1. A *group* is a set G admitting a binary multiplication operation, denoted $g \cdot h$ for group elements $g, h \in G$, which is subject to the following axioms:

(*a*) *Associativity*: $g \cdot (h \cdot k) = (g \cdot h) \cdot k$ for $g, h, k \in G$.

(*b*) *Identity*: The group contains a distinguished identity element, denoted e, satisfying $e \cdot g = g = g \cdot e$ for all $g \in G$.

(*c*) *Invertibility*: Each group element g has an inverse $g^{-1} \in G$ satisfying $g \cdot g^{-1} = g^{-1} \cdot g = e$.

Example 3.2. The simplest example of a group is the set $\mathbb{R}$ of real numbers, with addition being the group operation. The identity element is 0, and the inverse of x is its negative $-x$. Both the set of nonzero real numbers $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$ and the set of positive reals $\mathbb{R}^+$ form groups when the group operation is given by multiplication. The identity is the number 1, and inversion means taking reciprocals. All three groups are commutative, so $g \cdot h = h \cdot g$ for all group elements g, h. In group theory, commutative groups are called *abelian* in honor of Abel.

Example 3.3. The set of all invertible $n \times n$ real matrices forms a group, known as the (real) general linear group, and denoted $\mathrm{GL}(n, \mathbb{R})$. The group operation is matrix multiplication, and the identity element is the identity matrix; matrix inversion defines the inverse. Except in the case $n = 1$, which corresponds to the multiplicative group $\mathbb{R}^*$, the general linear group $\mathrm{GL}(n, \mathbb{R})$ forms a non-abelian group. Analogously, the complex general linear group $\mathrm{GL}(n, \mathbb{C})$ consists of all invertible $n \times n$ complex matrices. We will, at times, employ the abbreviated notation $\mathrm{GL}(n)$ to mean either the real or complex general linear group — the precise version will either be irrelevant or clear from the context.

Exercise 3.4. Let G and H be groups. Show how their Cartesian product $G \times H$ can be naturally endowed with the structure of a group.

Definition 3.5. A subset $H \subset G$ of a group G forms a *subgroup* provided the group operations on G define a group structure on H.

Example 3.6. The set of integers $\mathbb{Z} \subset \mathbb{R}$ forms a subgroup of the additive group of real numbers. According to Exercise 3.4, the vector space $\mathbb{R}^n$ forms a group with matrix addition defining the group operation. The set $\mathbb{Z}^n$ of integral vectors forms a discrete subgroup.

Exercise 3.7. Suppose G is a group. Prove that $H \subset G$ is a subgroup if and only if it is closed under the group operations, meaning that if $h, k \in H$, then $h \cdot k \in H$ and $h^{-1} \in H$.

Example 3.8. One of the most important subgroups of the general linear group $\mathrm{GL}(n, \mathbb{R})$ is the *special linear group*

$$\mathrm{SL}(n, \mathbb{R}) = \{A \mid \det A = 1\}, \tag{3.1}$$

consisting of all unimodular (unit determinant) matrices. It forms a subgroup because (a) the determinant of the product of two matrices equals the product of their determinants, and hence the product of two unimodular matrices is unimodular, and (b) the determinant of the inverse of a matrix is the reciprocal of its determinant, so that the inverse of a unimodular matrix is unimodular. In general, subgroups of $\mathrm{GL}(n)$ are known as *matrix groups*.

Exercise 3.9. Prove that the set $\mathrm{SL}(n, \mathbb{Z})$ consisting of all unimodular $n \times n$ matrices having integer entries forms a subgroup of $\mathrm{GL}(n, \mathbb{R})$. On the other hand, show that the set $\mathrm{GL}(n, \mathbb{Z})$ of *all* $n \times n$ integer matrices is not a subgroup.

Exercise 3.10. Prove that the set of all nonzero real matrices of the form $\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$ forms a subgroup of $\mathrm{GL}(2, \mathbb{R})$. Show that the group operation coincides with that of the multiplicative group $\mathbb{C}^* = \mathbb{C} \setminus \{0\}$ consisting of all nonzero complex numbers $\alpha + i\beta$. The *circle group*

$$S^1 = \{\, e^{i\theta} = \cos\theta + i \sin\theta \,\} \subset \mathbb{C}^* \tag{3.2}$$

containing all complex numbers of unit modulus forms a subgroup of $\mathbb{C}^*$. Its counterpart in $\mathrm{GL}(2, \mathbb{R})$ is the subgroup $\mathrm{SO}(2)$ consisting of all planar rotations $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$; see Example 3.35.

Exercise 3.11. Suppose $H \subset G$ is a subgroup. Let $g \in G$ be a fixed element. Prove that the set $g \cdot H \cdot g^{-1} = \{ghg^{-1} \mid h \in H\}$ is also a subgroup, called the *conjugate subgroup* to H under g.

Definition 3.12. A subgroup $H \subset G$ is called *normal* if it equals its conjugate subgroups, so $gHg^{-1} = H$ for all $g \in G$.

Example 3.13. The set $Z = \{\lambda \mathbb{1} \,|\, 0 \neq \lambda \in \mathbb{R}\}$ consisting of all nonzero scalar multiples of the identity matrix is a normal subgroup of $\mathrm{GL}(n, \mathbb{R})$. This example is a special case of the following basic result.

Exercise 3.14. Let G be a group. The *center* $Z \subset G$ is the subset consisting of all group elements which commute with every element in G. Thus, $z \in Z$ if and only if $z \cdot g = g \cdot z$ for all $g \in G$. In particular, $Z = G$ if and only if G is abelian. Prove that the center of a group is a subgroup and, in fact, a normal subgroup. Is every normal subgroup contained in the center?

Group Homomorphisms

In the foundations of group theory, the maps that respect the group operations play a distinguished role. These are the "morphisms" of the category of groups.

Definition 3.15. A map $\rho : G \to H$ between groups G and H is called a *group homomorphism* if it satisfies

$$\rho(g \cdot h) = \rho(g) \cdot \rho(h), \qquad \rho(e) = e, \qquad \rho(g^{-1}) = \rho(g)^{-1}, \tag{3.3}$$

for all $g, h \in G$.

Exercise 3.16. Prove that the image $\rho(G) \subset H$ of a group homomorphism forms a subgroup of the target group H.

A group homomorphism $\rho : G \to H$ is called a *group isomorphism* if it is one-to-one and onto, in which case G and H are isomorphic (meaning identical) groups. More generally, if ρ is one-to-one, then it is called a group *monomorphism*, in which case its image $\rho(G) \subset H$ is a subgroup which is isomorphic to G itself.

For example, the map $\rho : \mathbb{C}^* \to \mathrm{GL}(2, \mathbb{R})$ that takes a complex number $\alpha + i\beta$ to the 2×2 matrix introduced in Exercise 3.10 defines a group monomorphism, which restricts to an isomorphism $\rho : S^1 \xrightarrow{\sim} \mathrm{SO}(2)$ between the circle group and the group of planar rotations.

Exercise 3.17. Show that any map $\rho : G \to H$ which satisfies the first two properties in (3.3) automatically preserves the inverse and hence defines a group homomorphism.

Example 3.18. The map $t \mapsto e^t$ defines a group isomorphism $\rho : \mathbb{R} \to \mathbb{R}^+$ from the additive group of real numbers to the multiplicative

group of positive real numbers. The particular maps

$$\rho_1(t) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \quad \rho_2(t) = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \quad \rho_3(t) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix},$$

define different monomorphisms from $\mathbb{R}$ to the general linear group $\mathrm{GL}(2)$. All of these examples are special cases of the general matrix exponential function $\rho_J(t) = e^{tJ}$, where J is a fixed $n \times n$ matrix. The fact that $\rho_J : \mathbb{R} \to \mathrm{GL}(n)$ defines a group homomorphism is an immediate consequence of the usual properties of the matrix exponential:

$$e^{(t+s)J} = e^{tJ} \cdot e^{sJ}, \qquad e^{0J} = \mathbb{1}, \qquad e^{-tJ} = \left(e^{tJ}\right)^{-1}. \tag{3.4}$$

Thus, the image $H_J = \{\exp tJ\}$ of ρ_J forms an abelian subgroup $\mathrm{GL}(n)$. These "one-parameter subgroups" play an extremely important role in the theory of Lie groups and will be discussed in detail in Chapter 9.

Exercise 3.19. Prove that the *contragredient* map $\rho_c : A \mapsto A^{-T}$ defines a group isomorphism $\rho_c : \mathrm{GL}(n) \xrightarrow{\sim} \mathrm{GL}(n)$.

If G is a group, and $H \subset G$ a subgroup, then the *quotient space* G/H is defined as the set of all left cosets $g \cdot H = \{g \cdot h \,|\, h \in H\}$ for each $g \in G$. In general, G/H does not carry any natural group structure. Let $\pi : G \to G/H$ be the natural projection that maps a group element g to its coset $g \cdot H$.

Proposition 3.20. *If $H \subset G$ is a normal subgroup, then the quotient space G/H can be naturally endowed with the structure of a group such that the projection $\pi : G \to G/H$ is a group homomorphism.*

Proof: To make π into a group homomorphism, we should define the group operations on G/H so that $\pi(g) \cdot \pi(\hat{g}) = \pi(g \cdot \hat{g})$ and $\pi(g^{-1}) = \pi(g)^{-1}$ for any $g, \hat{g} \in G$. The identity element in G/H will correspond to the identity coset $\pi(e) = H$. We need to show that the product is well defined. Two group elements g and g' will map to the same coset, $\pi(g) = \pi(g')$, if and only if $g' = gh$ for some $h \in H$. Suppose $\hat{g}$ and $\hat{g}'$ also map to the same coset, $\pi(\hat{g}) = \pi(\hat{g}')$ so $\hat{g}' = \hat{g}\hat{h}$ for some $\hat{h} \in H$. Then $g'\,\hat{g}' = g\hat{g}(\hat{g}^{-1}h\hat{g})\hat{h}$. Now, since H is a normal subgroup, $\hat{g}^{-1}h\hat{g} = \tilde{h}$ is also an element of H. Thus, $g'\,\hat{g}' = gh\hat{g}\hat{h} = g\hat{g}\tilde{h}\hat{h} = (g\hat{g})\bar{h}$, where $\bar{h} = \tilde{h}\hat{h} \in H$ also. This implies that $g \cdot \hat{g}$ and $g' \cdot \hat{g}'$ lie in the same coset, proving that the induced group multiplication on G/H is well defined. The fact that the group inversion on G/H is also well defined follows by a similar computation. *Q.E.D.*

Exercise 3.21. Show that the set $2\pi\mathbb{Z} = \{0, \pm 2\pi, \pm 4\pi, \ldots\}$ of integer multiples of 2π forms a normal subgroup of $\mathbb{R}$; moreover the quotient group $\mathbb{R}/2\pi\mathbb{Z} \simeq S^1$ is isomorphic to the circle group. In a similar vein, show that $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$ forms an abelian group containing n elements. The group operation in $\mathbb{Z}_n$ is "addition mod n".

We next describe a more refined version of the result in Exercise 3.16, which provides a group-theoretic counterpart to the fundamental theorem for linear transformations between vector spaces.

Theorem 3.22. *Let $\rho: G \to H$ be a group homomorphism. The image $\rho(G)$ forms a subgroup of H. The kernel $K = \{k \in G \,|\, \rho(k) = e\}$ forms a normal subgroup of G. Moreover, the quotient group G/K is naturally isomorphic to the image $\rho(G)$ under the induced group monomorphism $\widetilde{\rho}: G/K \to H$.*

Proof: First, to show K is normal, for any $k \in K$, $g \in G$, we have

$$\rho(gkg^{-1}) = \rho(g)\rho(k)\rho(g^{-1}) = \rho(g) \cdot e \cdot \rho(g)^{-1} = e,$$

and hence $gkg^{-1} \in K$. We define $\widetilde{\rho}$ as in the statement of the theorem, so that $\widetilde{\rho}(\pi(g)) = \rho(g)$ for $g \in G$. This is well defined since

$$\widetilde{\rho}(\pi(g \cdot k)) = \rho(g \cdot k) = \rho(g)\rho(k) = \rho(g)e = \widetilde{\rho}(\pi(g)).$$

Finally, the fact that $\widetilde{\rho}$ is a monomorphism follows immediately from the definition of K. *Q.E.D.*

Corollary 3.23. *A group homomorphism $\rho: G \to H$ forms a monomorphism if and only if $\ker \rho = \{e\}$, that is, the only element of G mapped to the identity element of H is the identity $e \in G$.*

Example 3.24. The quotient group of the general linear group $\mathrm{GL}(n, \mathbb{R})$ by its center $Z = \{\lambda \mathbb{1}\}$ is known as the *projective linear group* and denoted by $\mathrm{PSL}(n, \mathbb{R}) = \mathrm{GL}(n, \mathbb{R})/\{\lambda \mathbb{1}\}$. This group plays the underlying role in the geometry of real projective space. If n is odd, then we can identify $\mathrm{PSL}(n, \mathbb{R}) \simeq \mathrm{SL}(n, \mathbb{R})$ with the special linear group. This follows from Theorem 3.22 if we use the group homomorphism $\rho(A) = (\det A)^{-1/n} A$, which maps $\mathrm{GL}(n, \mathbb{R})$ to $\mathrm{SL}(n, \mathbb{R})$. (We are using the fact that, for n odd, each real number has a unique real n^{th} root.) For n even, this is not quite correct; we can identify $\mathrm{PSL}(n, \mathbb{R}) \simeq \mathrm{SL}(n, \mathbb{R})/\{\pm \mathbb{1}\}$, so that $\mathrm{SL}(n, \mathbb{R})$ forms a two-fold covering of the projective group.

Exercise 3.25. Motivated by (1.12) (although the reader should

note that the matrix has been transposed here), prove that the map

$$\rho\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \alpha^2 & 2\alpha\beta & \beta^2 \\ \alpha\gamma & \alpha\delta + \beta\gamma & \beta\delta \\ \gamma^2 & 2\gamma\delta & \delta^2 \end{pmatrix} \tag{3.5}$$

forms a group homomorphism $\rho\colon \mathrm{GL}(2) \to \mathrm{GL}(3)$. Determine the kernel.

Exercise 3.26. For which matrices J does the matrix exponential $\rho_J(t) = e^{tJ}$ form a group monomorphism? (*Hint*: Look at the Jordan canonical form of J.) If ρ_J is not a monomorphism, then either $J = 0$ or the image $\rho_J(\mathbb{R})$ is isomorphic to the circle group S^1.

Transformation Groups

Although the abstract theory of groups is well worth developing for its own intrinsic beauty, the real power of the concept is only revealed when the group acts on some space. Indeed, in the last century, groups per se did not exist in the abstract, as is now standard, but always arose concretely through their action as groups of transformations. Such "transformation groups" all arise as subgroups of the following general example.

Example 3.27. Let X be any set. Let $\mathcal{G} = \mathcal{G}(X)$ denote the set of all one-to-one maps $\varphi\colon X \to X$. Then $\mathcal{G}$ forms a group in which the group operation is defined by composition of maps. The identity transformation $\mathbb{1}_X$ plays the role of the group identity element, and the inverse of a transformation is defined as the usual functional inverse. The basic properties of functional composition and inversion automatically imply that $\mathcal{G}$ satisfies the group axioms in Definition 3.1.

Example 3.28. Let X be a finite set with $n = \#X$ elements. Any invertible transformation $\varphi\colon X \to X$ induces a permutation of the elements of X, and hence we can identify $\mathcal{G}(X)$ with the group of permutations of n objects. Thus, $\mathcal{G}(X)$ forms a finite group having $n!$ elements, known as the *symmetric group* on n objects, and denoted $\mathbb{S}^n$. For example, if $X = \{1, 2, 3\}$ has three elements, labeled by the numbers from 1 to 3, then $\mathcal{G}(X) = \mathbb{S}^3$ consists of the $6 = 3!$ permutations

$$\mathbb{S}^3 = \Big\{ (123), (132), (213), (231), (312), (321) \Big\}, \tag{3.6}$$

where (ijk) denotes the permutation that maps 1 to i, 2 to j, 3 to k. Note that $\mathbb{S}^n$ is a non-abelian group when $n \geq 3$, while $\mathbb{S}^2 \simeq \mathbb{Z}_2$.

If the space X comes equipped with additional structure, then one might impose corresponding restrictions on the class of allowable maps, leading to important subgroups of the vast group $\mathcal{G}(X)$. The main requirement is that the relevant constraints on the maps must be preserved under composition and inversion.

Example 3.29. If $X = \mathbb{R}^n$, then the allowable transformations in $\mathcal{G}(X)$ are usually required to satisfy topological or differentiability constraints. For example, the subgroup $\mathcal{C}^0(X)$ consists of all continuous, invertible maps $\varphi\colon X \to X$. Here we are using the fact that the composition of two continuous functions is continuous, as is the inverse of a continuous one-to-one map. Similarly, $\mathcal{C}^k(X)$, where $0 \le k \le \infty$, is defined to be the group of all invertible continuously k times differentiable maps — *diffeomorphisms*. Even more restrictively, the analytic structure of X serves to define the subgroup $\mathcal{A}(X)$ of *analytic diffeomorphisms*. This example can be readily extended to the case when X is an analytic manifold, e.g., a surface, as defined in Chapter 8.

Definition 3.30. A *transformation group* acting on a space X is defined by a group homomorphism $\rho\colon G \to \mathcal{G}(X)$ mapping a given group G to the group of invertible maps on X.

In other words, each element $g \in G$ will induce an invertible map $\rho(g)\colon X \to X$. In order that this identification define a group homomorphism, we must require that ρ satisfy the basic properties

$$\rho(g \cdot h) = \rho(g) \circ \rho(h), \qquad \rho(e) = \mathbb{1}_X, \qquad \rho(g^{-1}) = \rho(g)^{-1}, \tag{3.7}$$

for each $g, h \in G$. According to Exercise 3.7, any set $\widetilde{G} \subset \mathcal{G}(X)$ consisting of invertible maps $\varphi\colon X \to X$ which is closed under composition $\varphi \circ \psi$ and inversion φ^{-1} will form a subgroup of $\mathcal{G}(X)$ and hence determines a transformation group on X. In all the examples that will be considered in this book, the space X carries an analytic structure, and the transformation group $\rho\colon G \to \mathcal{A}(X)$ consists of analytic diffeomorphisms.

For a fixed group action, it is common to write $g \cdot x$ for the action of the group element $g \in G$ on the point $x \in X$, instead of the more cumbersome notation $\rho(g)(x)$. Thus, conditions (3.7) become

$$g \cdot (h \cdot x) = (g \cdot h) \cdot x, \qquad e \cdot x = x, \qquad g \cdot (g^{-1} \cdot x) = x, \tag{3.8}$$

for all $g, h \in G$, $x \in X$. In the first condition, $h \cdot x = \rho(h)(x)$ denotes the action of h on x, whereas $g \cdot h$ denotes the group multiplication, which can be identified with composition $\rho(g) \circ \rho(h)$ between the corresponding transformations.

Example 3.31. Any group acts on itself by left multiplication. In other words, given G, we set $X = G$ also and let $\lambda: G \to \mathcal{G}(G)$ map the group element g to the left multiplication map $\lambda(g): h \mapsto g \cdot h$. Note that this action is compatible with our notational convention (3.8). A closely related action is given by right multiplication $\rho(g): h \mapsto h \cdot g^{-1}$ acting on $X = G$. The inverse is required so that ρ forms a group homomorphism: $\rho(g \cdot h) = \rho(g) \cdot \rho(h)$.

Exercise 3.32. Define the notions of homomorphism and isomorphism for transformation group actions. When are the left and right actions of a group on itself isomorphic?

Example 3.33. Let $X = V$ be a real vector space. Consider the group $\mathrm{GL}(V) \subset \mathcal{G}(V)$ consisting of all invertible linear transformations $T: V \to V$, i.e., one-to-one maps that satisfy $T(x+y) = T(x) + T(y)$, for $x, y \in V$, and $T(\lambda x) = \lambda T(x)$ for λ a scalar. If V is finite-dimensional, then we can introduce a basis $\{e_1, \dots, e_n\}$ so as to identify $V \simeq \mathbb{R}^n$ and thereby identify each invertible linear transformation with its $n \times n$ matrix representative. In this manner, the abstract general linear group $\mathrm{GL}(n, \mathbb{R})$ introduced in Example 3.3 is naturally realized as the group of all invertible linear transformations on an n-dimensional vector space, where composition serves to define the group operation. Note that any matrix subgroup of $\mathrm{GL}(n, \mathbb{R})$ also acts on $\mathbb{R}^n$ via linear transformations. These are the linear group actions or representations, destined to play the pivotal role in classical invariant theory.

Example 3.34. A smooth transformation $\varphi: \mathbb{R}^n \to \mathbb{R}^n$ is called *orientation-preserving* if its Jacobian matrix has positive determinant everywhere. The set of invertible orientation-preserving transformations forms a subgroup of $\mathcal{C}^k(\mathbb{R}^n)$ for $k \geq 1$. In particular the group of linear, orientation-preserving transformations is the subgroup $\mathrm{GL}(n, \mathbb{R})^+ = \{\det A > 0\}$ consisting of $n \times n$ matrices with positive determinant. A transformation φ is called *volume-preserving* if $\mathrm{vol}\, \varphi(S) = \mathrm{vol}\, S$ for every subset $S \subset \mathbb{R}^n$, where vol denotes the ordinary Lebesgue measure. The special linear group $\mathrm{SL}(n, \mathbb{R})$, (3.1), consists of linear transformations that preserve both volume and orientation.

Example 3.35. We denote the usual Euclidean norm on $\mathbb{R}^n$ by $\| x \| = \sqrt{(x_1)^2 + \cdots + (x_n)^2}$. A transformation $\varphi: \mathbb{R}^n \to \mathbb{R}^n$ is called an *isometry* if $\| \varphi(x) - \varphi(y) \| = \| x - y \|$ for all $x, y \in \mathbb{R}^n$, i.e., it preserves (Euclidean) distances. The group of linear isometries forms a subgroup $\mathrm{O}(n) \subset \mathrm{GL}(n, \mathbb{R})$ of the general linear group, known as the *orthogonal*

group; it can be characterized as $\mathrm{O}(n) = \{A \in \mathrm{GL}(n,\mathbb{R}) \,|\, A^T A = \mathbb{1}\}$. The orthogonal group contains both rotations and reflections, distinguished by the sign of their determinant. The *rotation* or *special orthogonal group* $\mathrm{SO}(n) = \{A \in \mathrm{O}(n) \,|\, \det A = +1\}$ consists of all orientation-preserving orthogonal transformations.

Example 3.36. Slightly generalizing Example 3.33, we recall that an *affine transformation* of the linear space $\mathbb{R}^n$ is a combination of a linear transformation and a translation, and hence has the general form $x \mapsto A\,x + a$, where $A \in \mathrm{GL}(n,\mathbb{R})$ is an invertible matrix and $a \in \mathbb{R}^n$ a fixed vector. The composition of two affine transformations is also affine, as is the inverse. Therefore the set $\mathrm{A}(n) = \mathrm{A}(n,\mathbb{R})$ of all affine transformations forms a group — the *affine group*— which is parametrized by the pair $(A,a) \in \mathrm{GL}(n,\mathbb{R}) \times \mathbb{R}^n$. Although as a set $\mathrm{A}(n)$ can be identified with the Cartesian product of the groups $\mathrm{GL}(n,\mathbb{R})$ and $\mathbb{R}^n$, it is *not* isomorphic to the Cartesian product group $\mathrm{GL}(n,\mathbb{R}) \times \mathbb{R}^n$ because its group multiplication law, $(A,a)\cdot(B,b) = (AB, a + Ab)$, is *not* the same as the Cartesian product group action. This forms a particular case of a general construction known as the *semi-direct product*, cf. [**169**; p. 37], and often denoted by $\mathrm{A}(n) = \mathrm{GL}(n,\mathbb{R}) \ltimes \mathbb{R}^n$.

Exercise 3.37. Prove that the map $\rho(A,b) = \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}$ defines a group monomorphism $\rho\colon \mathrm{A}(n) \to \mathrm{GL}(n+1,\mathbb{R})$, realizing the affine group as a matrix group in one higher dimension.

Exercise 3.38. An affine transformation is called *equi-affine* if it preserves volume. Show that the set of equi-affine transformations forms a subgroup $\mathrm{SA}(n) \subset \mathrm{A}(n)$ of the affine group, which is defined by the unimodularity constraint $\det A = 1$. The *Euclidean group* is defined as the subgroup of $\mathrm{A}(n)$ whose linear part is orthogonal, so

$$\mathrm{E}(n) = \{\, (R,a) \mid\ R \in \mathrm{O}(n), a \in \mathbb{R}^n \,\} = \mathrm{O}(n) \ltimes \mathbb{R}^n \subset \mathrm{A}(n).$$

Prove that the Euclidean group is the group of affine isometries. In fact, it can be shown that every isometry is necessarily affine, cf. [**240**; §2.3], and thus the Euclidean group is the full isometry group of Euclidean space. As such, it plays the foundational role in Euclidean geometry. The *proper Euclidean group* $\mathrm{SE}(n) = \mathrm{E}(n) \cap \mathrm{SA}(n) \simeq \mathrm{SO}(n) \ltimes \mathbb{R}^n$ consists of the orientation-preserving Euclidean transformations.

Given a transformation group action defined by a homomorphism $\rho\colon G \to \mathcal{G}(X)$, the image $\rho(G)$ will form a subgroup of the group of all

invertible maps of the space X. The kernel

$$K = \{g \in G \mid \rho(g) = \mathbb{1}_X\} = \{g \mid g \cdot x = x \text{ for all } x \in X\}$$

is known as the *global isotropy subgroup*. It forms a normal subgroup of the transformation group G and consists of all group elements which act completely trivially on the space. According to Theorem 3.22, we can identify $\rho(G)$ with the quotient group G/K.

Definition 3.39. A transformation group is said to act *effectively* or *faithfully* if it has trivial global isotropy subgroup: $K = \{e\}$.

The condition means that the only group element acting as the identity transformation on X is the identity element of G. A group acts effectively if and only if the map $\rho: G \to \mathcal{G}(X)$ is a monomorphism, which means that different group elements have different effects: $g \cdot x = h \cdot x$ for all $x \in X$ if and only if $g = h$. Theorem 3.22 shows that any non-effective transformation group G can, without any significant loss of information or generality, be replaced by the effectively acting quotient group $\widehat{G} = G/K$.

Proposition 3.40. *Suppose G is a transformation group acting on a space X, and let K denote the global isotropy subgroup. There is a well defined effective action of the quotient group $\widehat{G} = G/K$ on X, which "coincides" with that of G in the sense that two group elements g and $\tilde{g}$ have the same effect on X, so $g \cdot x = \tilde{g} \cdot x$ for all $x \in X$, if and only if they have the same image in $\widehat{G}$, so $\tilde{g} = g \cdot k$ for some $k \in K$.*

Example 3.41. The linear fractional transformations

$$A \cdot p = \frac{\alpha p + \beta}{\gamma p + \delta}, \qquad A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}(2), \tag{3.9}$$

define an action of the general linear group $\mathrm{GL}(2, \mathbb{R})$ on the real projective line $\mathbb{RP}^1$. The action is *not* effective since multiples of the identity matrix act trivially. The global isotropy subgroup is $K = \{\lambda \mathbb{1}\}$, and the effectively acting quotient group given in Proposition 3.40 is the projective linear group $\mathrm{PSL}(2, \mathbb{R}) = \mathrm{GL}(2, \mathbb{R})/\{\lambda \mathbb{1}\}$ defined in Example 3.24. This example can be straightforwardly generalized to real and complex projective actions in higher dimensions.

Symmetry Groups, Invariant Sets, and Orbits

As remarked earlier, groups originally arose because they effectively crystallized the intuitive notion of symmetry and as such lie at the founda-

tions of geometry, physics, art, and human perception. In general, a transformation group is a symmetry group of an object if its action leaves the object unchanged.

Definition 3.42. Let $Y \subset X$. A *symmetry* of Y is an invertible transformation $\varphi\colon X \to X$ that leaves Y fixed, so $\varphi(Y) = Y$.

The crucial observation, dating back to Lagrange and Galois, is that the collection of all symmetries of a subset $Y \subset X$ forms a subgroup $\mathcal{S}(Y) \subset \mathcal{G}(X)$ of the group of all invertible transformations on X, known as the *symmetry group* of Y. Indeed, the identity transformation $\mathbb{1}_X \in \mathcal{S}(Y)$ is clearly always a symmetry; further, if $\varphi(Y) = Y$ and $\psi(Y) = Y$, then $\psi \circ \varphi(Y) = \psi(\varphi(Y)) = Y$, and $\varphi^{-1}(Y) = Y$.

Remark: If we relax the original assumption and require only that $\varphi(Y) \subset Y$, then the inversion property does not hold and the set of such maps generally only forms a "semi-group". For example, if $Y = (-1, 1) \subset \mathbb{R}$, then the transformation $\varphi(x) = \frac{1}{2}x$ satisfies $\varphi(Y) \subset Y$, whereas $\varphi^{-1}(x) = 2x$ clearly does not.

Generally, one does not deal with the entire symmetry group of a given subset but rather admits only those symmetries satisfying suitable constraints — for example, continuous symmetries, analytic symmetries, linear symmetries, and projective symmetries. In other words, starting with a given transformation group $\rho\colon G \to \mathcal{G}(X)$, the *symmetry subgroup* of a subset $Y \subset X$ is $G_Y = \rho^{-1}(\mathcal{S}(Y)) = \{g \in G \,|\, g \cdot Y = Y\}$. A transformation group G acting on X is said to be a *symmetry group* of the subset Y if *every* group element is a symmetry, so that $G_Y = G$. In this case, Y is designated a *G–invariant subset* of X.

Example 3.43. Consider the planar equilateral triangle with vertices $\Delta = \left\{(1,0), (-\frac{1}{2}, \frac{\sqrt{3}}{2}), (-\frac{1}{2}, -\frac{\sqrt{3}}{2})\right\} \subset \mathbb{R}^2$. A linear transformation on $\mathbb{R}^2$ defines a symmetry of Δ if and only if it has one of the following six matrix representatives:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}. \tag{3.10}$$

Consequently, the linear symmetry group of an equilateral triangle is isomorphic to the symmetric group $\mathbb{S}^3$. On the other hand, the isosceles

triangle with vertices $\widetilde{\Delta} = \{(1,0),(0,1),(0,-1)\}$ has only two linear symmetries: the identity and the reflection $y \mapsto -y$ through the x-axis. If we enlarge our class of transformations to include affine maps, then the isosceles triangle $\widetilde{\Delta}$ admits six affine symmetries, which we represent as 3×3 matrices as in Exercise 3.37:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix},$$

$$\begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{3}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{3}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}.$$

Indeed, the affine symmetry group of *any* triangle is isomorphic to $\mathbb{S}^3$.

For the square with vertices $S = \{(1,0),(0,1),(-1,0),(0,-1)\}$, the linear and affine symmetry groups coincide and consist of eight linear isometries: the identity; rotations through 90°, 180°, and 270°; reflections through the two coordinate axes; and reflections through the two lines making 45° angles with the axes. The reader should verify that these transformations form a non-abelian group, known as the *dihedral group* D_4. Note that there are no linear (or affine) symmetries realizing every possible permutation of the four vertices; for instance, we cannot leave two adjacent vertices fixed and interchange the other two. There are, however, nonlinear transformations that realize such permutations. For example, the projective transformation†

$$(x,y) \longmapsto \left(\frac{x-y+1}{2(x+y)}, \frac{-x+y+1}{2(x+y)} \right)$$

fixes $(1,0)$ and $(0,1)$, while interchanging $(-1,0)$ and $(0,-1)$. (Since projective transformations map lines to lines, they also preserve the edges of the square.) In fact, there are 24 projective symmetries of a square, which form a group isomorphic to $\mathbb{S}^4$, realized as a group of 3×3 matrices.

Exercise 3.44. Prove that a generic quadrilateral has no nontrivial affine symmetries but always has 24 projective symmetries. Show

† See Chapter 10 for details.

that a triangle has an infinite number of projective symmetries. What is its isotropy group, that is, the subgroup leaving all three vertices fixed? Investigate the linear, affine, and projective symmetries of pentagons.

Given a transformation group G acting on a space X, the symmetry group of a single point $x \in X$ is known as its *isotropy subgroup*: $G_x = \{g \in G \,|\, g \cdot x = x\}$. The transformation group acts *freely* if all isotropy subgroups are all trivial: $G_x = \{e\}$ for all $x \in X$. This is equivalent to the statement that $g \cdot x = h \cdot x$ for any one point $x \in X$ if and only if $g = h$. Free actions should be contrasted with effective actions, where $g = h$ if and only if $g \cdot x = h \cdot x$ for *all* $x \in X$. For example, the rotation group $\mathrm{SO}(3)$ acts effectively on three-dimensional space $X = \mathbb{R}^3$, since the only rotation which leaves every point fixed is the identity; however, it does not act freely, since any nonzero point $0 \neq x \in \mathbb{R}^3$ is fixed by the rotations around the axis formed by the line passing through x and 0. And, of course, the origin is left fixed by all rotations. Similar remarks hold for $\mathrm{SO}(n)$ acting on $\mathbb{R}^n$ provided $n \geq 3$.

Given a transformation group G, we would like to characterize the possible G–invariant subsets. Trivially, X itself is invariant. Technically speaking, the empty set $\varnothing \subset X$ is also G–invariant, but this is uninteresting. Note that unions and intersections of G–invariant subsets are also G–invariant. The minimal G–invariant subsets are of particular importance.

Definition 3.45. An *orbit* of a transformation group is a minimal nonempty invariant subset. In particular, a *fixed point* is a G–invariant point $x_0 \in X$, so that $g \cdot x_0 = x_0$ for all $g \in G$.

Example 3.46. As a simple example, consider the standard action

$$(x, y) \longmapsto (x \cos\theta - y \sin\theta, x \sin\theta + y \cos\theta) \tag{3.11}$$

of the rotation group $\mathrm{SO}(2)$ on $\mathbb{R}^2$. Any circle $\{x^2 + y^2 = r^2\}$ centered at the origin is a rotationally invariant subset of the plane. Since the circles are minimal — they contain no nonempty rotationally invariant subset — they are the orbits of $\mathrm{SO}(2)$. The only fixed point is the origin. Every other invariant subset, e.g., an annulus $\{a < x^2 + y^2 < b\}$, is necessarily a union of circles.

Example 3.47. For the usual linear action of $\mathrm{GL}(n)$ on $\mathbb{R}^n$, there are two orbits: the origin $\{0\}$ and the remainder $\mathbb{R}^n \setminus \{0\}$. The same holds for $\mathrm{SL}(n)$ since we can still map any nonzero vector in $\mathbb{R}^n$ to any other nonzero vector by a matrix of determinant 1.

Proposition 3.48. *Given a transformation group acting on a space X, the orbit $\mathcal{O}_x$ through a point $x \in X$ is just the set of all images of x under arbitrary group transformations: $\mathcal{O}_x = \{g \cdot x \,|\, g \in G\}$. A subset $S \subset M$ is G–invariant if and only if it is the union of orbits.*

Exercise 3.49. Let $x, y \in X$ be points lying in the same orbit of G. Prove that their isotropy subgroups G_x and G_y are conjugate subgroups of G. More generally, given $Y \subset X$, show that the symmetry group of $g \cdot Y = \{g \cdot y \,|\, y \in Y\}$ is conjugate to that of Y.

A group action is called *transitive* if there is only one orbit, so for every $x, y \in X$ there exists at least one $g \in G$ such that $g \cdot x = y$. In this case, the only G–invariant subsets are the trivial ones $\varnothing$ and X. For example, the linear actions of the groups $\mathrm{GL}(n, \mathbb{R})$, $\mathrm{SL}(n, \mathbb{R})$, and $\mathrm{SO}(n)$ on $\mathbb{R}^n$ all induce transitive projective actions on the space $\mathbb{RP}^{n-1}$.

Exercise 3.50. Is the action of the real rotation group $\mathrm{SO}(2)$ on the complex projective plane $\mathbb{CP}^1$ given by "linear fractional rotations" $p \mapsto (p \cos\theta - \sin\theta)/(p \sin\theta + \cos\theta)$ transitive? effective? free? Describe the orbits.

Exercise 3.51. Prove that if G acts freely and transitively on X, then we can naturally identify $X \simeq G$, and the action is isomorphic to the left multiplication action of Example 3.31.

Equivalence and Canonical Forms

Given a transformation group G acting on X, we shall call two points $x, y \in X$ *equivalent* if there exists a group transformation mapping one to the other, so that $y = g \cdot x$ for some $g \in G$. In other words, two points are equivalent if and only if they lie in the same orbit. The *equivalence problem* for a transformation group is to find necessary and sufficient conditions for this to hold. A trivial case is when the group acts transitively, which implies that all points are equivalent. The solution to an equivalence problem typically requires the construction of suitable invariants which serve to distinguish the orbits of the group.

In this context, a *canonical form* of an element $x \in X$ just means a distinguished, "simple" representative $x_0 \in \mathcal{O}_x$ of the orbit containing x. Thus, a complete list of canonical forms can be identified with a list of orbits of the group, since each orbit must contain one (and, in an irredundant list, only one) canonical form, which thereby serves to

distinguish the orbit. Of course, there is no uniquely specified canonical form, and some choice, usually based on one's æsthetic judgment of "simplicity", must be exercised. Note that the determination of a complete system of canonical forms leads to an immediate solution to the associated equivalence problem: Two objects are equivalent if and only if they have the same canonical form.

For example, the well known Jordan canonical form of a matrix corresponds to the conjugation action $X \mapsto AXA^{-1}$ of the complex general linear group $\mathrm{GL}(n, \mathbb{C})$ on the space of $n \times n$ complex matrices X. However, other less commonly known canonical forms, including the older "rational canonical form", cf. [**219**], can also be advantageously utilized.

The "coordinates" of the canonical form are particular invariants, often referred to as the *moduli* for the given transformation group, cf. [**156, 213**]. For the conjugation action of $\mathrm{GL}(n, \mathbb{C})$ on matrices, the moduli are the eigenvalues of the matrix and the sizes of the different Jordan blocks. Thus, moduli can be both continuous and discrete; in regular cases, as discussed in Chapter 8, the moduli provide a complete system of functionally independent invariants for the group action.

Example 3.52. For the rotation group $\mathrm{SO}(n)$ acting on $\mathbb{R}^n$, the orbits are the spheres $\{\| x \| = \text{constant}\}$ centered at the origin. Hence, two points in $\mathbb{R}^n$ are equivalent if and only if they have the same norm: $\| x \| = \| y \|$. Thus, we can choose the canonical form of a vector $x \in \mathbb{R}^n$ to be a non-negative multiple of the first unit basis vector, $r\, e_1$, with $r = \| x \| \geq 0$ being the single modulus for the group action. (Clearly, there is no particular reason to choose the first basis vector for the canonical form.) As for the actions of $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$, each vector $x \in \mathbb{R}^n$ has just two possible canonical forms: either 0 or, say, e_1.

Example 3.53. Consider the action of the group $\mathrm{GL}(2, \mathbb{R})$ of real invertible 2×2 matrices on the space $\mathcal{Q}$ of real quadratic forms analyzed in Chapter 1. Since each quadratic form is uniquely determined by its three coefficients, we can identify $\mathcal{Q} \simeq \mathbb{R}^3$. The induced action of $\mathrm{GL}(2, \mathbb{R})$ on $\mathcal{Q}$ is explicitly given in (1.12). What are the orbits? According to the classification in Chapter 1, there are four different canonical forms for real quadratic forms, and hence there are four distinct orbits. These are almost completely distinguished by the discriminant; the first orbit consists of all those quadratic forms with positive discriminant; the second consists of those with negative discriminant. There is a fixed point consisting of the zero quadratic form; the final orbit consists of

all not identically zero quadratic forms with zero discriminant. In terms of the coordinates (a, b, c) provided by the coefficients of the quadratic form (1.7), the orbits are

$$\mathcal{O}_+ = \{ac > b^2\}, \qquad \mathcal{O}_0 = \{ac = b^2,\ a^2 + b^2 + c^2 \neq 0\},$$
$$\mathcal{O}_- = \{ac < b^2\}, \qquad \mathcal{O}_* = \{a = b = c = 0\}.$$

The isotropy subgroup for a given quadratic form is its (linear) symmetry group and consists of all linear transformations that leave the form unchanged. The positive definite canonical form x^2+y^2 was already considered in Example 3.35 — its symmetry group is the orthogonal subgroup $\mathrm{O}(2) \subset \mathrm{GL}(2)$; the same holds true for the negative definite canonical form $-x^2-y^2$. The symmetry group of the indefinite canonical form x^2-y^2 consists of hyperbolic rotations $\begin{pmatrix} \pm\cosh t & \pm\sinh t \\ \pm\sinh t & \pm\cosh t \end{pmatrix}$, where the product of signs in each row of the matrix must be the same. The symmetry group of the alternative indefinite quadratic form xy is the conjugate subgroup containing all matrices $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$ and $\begin{pmatrix} 0 & \mu \\ \mu^{-1} & 0 \end{pmatrix}$, for $\lambda, \mu \neq 0$. One can distinguish the symmetry groups of definite versus indefinite forms by their topology: the symmetry group of a definite quadratic form is compact, while that of an indefinite form is not. As for the degenerate forms $\pm x^2$, they have the same symmetry group, consisting of all matrices of the form $\begin{pmatrix} \pm 1 & 0 \\ \gamma & \delta \end{pmatrix}$. Unlike the definite forms, the symmetry group depends on two parameters. Finally the trivial zero form has the entire four-parameter group $\mathrm{GL}(2)$ as a symmetry group. Thus, one can almost distinguish between the various orbits by the underlying structure of their isotropy subgroups.

Exercise 3.54. Use the canonical forms for cubic polynomials presented in Chapter 2 to determine the orbits of the action on $\mathrm{GL}(2, \mathbb{R})$ on the space of cubic forms. Determine the symmetry groups of the canonical cubic forms.

Exercise 3.55. Consider the induced action of the rotation group $\mathrm{SO}(2)$ on the space of quadratic forms $\mathcal{Q}$, obtained by restricting the general linear action to pure rotations. In other words, we use the transformation rules (1.12) with $\alpha = \delta = \cos\theta$, $-\beta = \gamma = \sin\theta$. Prove that a complete list of canonical forms is provided by the diagonal forms $Q_0(x, y) = \lambda x^2 + \mu y^2$. Is the list redundant, that is, can we map one diagonal quadratic form to another? Use this to describe the orbits.

The preceding examples are particular cases of the general theory of quadratic forms on finite-dimensional vector spaces. According to Sylvester's Law of Inertia, [**75**; §X.2, **219**; p. 89], the action $S \mapsto ASA^T$ of the general linear group $\mathrm{GL}(n, \mathbb{R})$ on the space of real symmetric $n \times n$ matrices S has a discrete set of orbits, which are classified by the matrix's *signature*, meaning the number of positive, zero, and negative eigenvalues. Canonical forms are provided by the diagonal matrices with $0 \leq p \leq n$ entries equal to $+1$, followed by $0 \leq q \leq n - p$ entries equal to -1, followed by $n - p - q$ zeros on the diagonal. In particular, the orbit containing the identity matrix consists of all symmetric, positive definite matrices.

Theorem 3.56. *Two symmetric matrices are equivalent under the action $S \mapsto ASA^T$, $A \in \mathrm{GL}(n, \mathbb{R})$, if and only if they have the same signature.*

If we restrict to the orthogonal subgroup $\mathrm{O}(n) \subset \mathrm{GL}(n, \mathbb{R})$, then we can still diagonalize any symmetric matrix S, leading to a canonical form $D = \mathrm{diag}(\lambda_1, \dots \lambda_n)$. Note that this canonical form is not quite uniquely determined, since we can rearrange the order of the entries λ_i. Thus, in this case, the moduli are provided by symmetric functions of the eigenvalues.

Remark: We can uniquely identify each symmetric matrix S with a quadratic form $Q(\mathbf{x}) = x^T S\, x$ on $\mathbb{R}^n$, and so Sylvester's Theorem and its orthogonal counterpart describe the canonical forms for multivariate quadratic forms under linear and orthogonal transformations.

Chapter 4

Representations and Invariants

While general transformation groups play a ubiquitous role in geometry, the most important for invariant theoretic purposes are the linear versions (along with their projective counterparts). Linear group actions are referred to as "representations", so as to distinguish them from the more general nonlinear transformation groups. Representation theory is a vast field of mathematical research, and we shall only have room to survey its most elementary aspects, concentrating on those which are relevant to our subject. More complete treatments can be found in numerous reference texts, including [**56**, **144**, **223**, **231**]. Representation theory has an amazing range of applications in physics and mathematics, including special function theory, [**152**, **212**, **225**], quantum mechanics and particle physics, [**46**, **136**, **232**, **238**], solid state physics, [**130**], probability, [**145**], harmonic and Fourier analysis, [**212**, **228**], number theory, [**79**, **118**, **145**], combinatorics, [**143**, **201**], bifurcation theory, [**82**, **190**], and many other fields. A quote of I. M. Gel'fand, "... all of mathematics is some kind of representation theory...", [**132**], is perhaps not as far-fetched as it might initially seem! The representations of a general (even nonlinear) transformation group on the scalar-valued functions defined on the space where the group acts are of particular importance. In classical invariant theory, our original transformation rules for inhomogeneous and homogeneous polynomials are very special cases of the general framework of group representations on function spaces. In this fashion, we are led back to our primary subject, renewed and reinvigorated by an understanding of the requisite mathematical theory.

Representations

For simplicity we shall state the basic concepts in representation theory in the context of real group actions on real vector spaces; the corresponding complex version is an immediate adaptation and it is left to the reader to fill in the details.

Definition 4.1. A *representation* of a group G is defined by a group homomorphism $\rho: G \to \mathrm{GL}(V)$ to the group of invertible linear transformations on a vector space V.

Thus, according to (3.3), any representation must satisfy the following basic rules:

$$\rho(g \cdot h) = \rho(g) \cdot \rho(h), \qquad \rho(e) = \mathbb{1}, \qquad \rho(g^{-1}) = \rho(g)^{-1}, \tag{4.1}$$

for all $g, h \in G$. The vector space on which the representation acts will be called the *representation space*; it is not necessarily finite-dimensional, although, in view of our particular applications, we can safely avoid the many analytical complications inherent in the infinite-dimensional context.

Example 4.2. Let $G = \mathbb{R}$ be the additive group of real numbers. According to (3.4), for any fixed $n \times n$ matrix J, the matrix exponential $\rho_J(t) = e^{tJ}$ defines an n-dimensional representation of $\mathbb{R}$. Particular cases appeared in Example 3.18.

Example 4.3. A well-known representation of the symmetric group $\mathbb{S}^n$ is provided by the $n \times n$ permutation matrices. To be specific, a permutation $\pi \in \mathbb{S}^n$ is represented by the linear transformation A_π which permutes the entries of vectors $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ accordingly: $A_\pi(x) = (x_{\pi(1)}, \dots, x_{\pi(n)})$. It is easy to see that the map $\rho: \pi \mapsto A_\pi$ defines a group monomorphism from $\mathbb{S}^n$ to $\mathrm{GL}(n)$ and hence determines a representation of the symmetric group. For example, the symmetric group $\mathbb{S}^3$, as given in (3.6), is represented by the following six matrices:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

See Example 3.43 for several other interesting representations of the symmetric group.

Example 4.4. Let $G = \mathrm{GL}(n, \mathbb{R})$ be the general linear group. The simplest possible representation is the trivial one-dimensional representation that assigns to each matrix $A \in \mathrm{GL}(n)$ the real number 1. Slightly more interesting is the one-dimensional determinantal representation

$\rho_d(A) = \det A$. More generally, any power $\rho_d^k(A) = (\det A)^k$ of the determinant also forms a one-dimensional representation of $\mathrm{GL}(n)$. An evident n-dimensional representation is provided by the identity representation $\rho_e(A) = A$ acting on $V = \mathbb{R}^n$. A second n-dimensional representation is the so-called *contragredient* or *dual* representation $\rho_c(A) = A^{-T}$ considered in Exercise 3.19. Higher dimensional representations can be constructed by a variety of techniques. For example, the conjugation action $X \mapsto AXA^{-1}$ defines a representation of $\mathrm{GL}(n)$ on the space $\mathcal{M}_{n\times n} \simeq \mathbb{R}^{n^2}$ of $n \times n$ matrices, as does the action $X \mapsto AXA^T$ arising in the theory of quadratic forms.

The preceding representations of the general linear group are particular instances of tensorial operations that can be applied to arbitrary representations. These include the operations of duality, direct sum, tensor product, and symmetric product. The simplest case is the direct sum $\rho \oplus \sigma$ of two representations ρ, σ of G on V, W, respectively, which acts in the obvious fashion on the sum $V \oplus W$ of the representation spaces.

If V is a real, finite-dimensional vector space, we let V^* denote the dual vector space, which is defined as the space of real-valued linear maps $\omega\colon V \to \mathbb{R}$. It is not difficult to see that V^* is a vector space of the same dimension as V. In terms of a basis $\{e_1, \dots, e_n\}$ of V, the corresponding dual basis $\{\varepsilon_1, \dots, \varepsilon_n\}$ of V^* is defined so that $\varepsilon_i(e_j) = \delta^i_j$, where δ^i_j is the usual Kronecker delta, which has the value 1 if $i = j$ and 0 otherwise. The dual of V^* can be identified with V again: $(V^*)^* \simeq V$; the identification takes $v \in V$ to the linear map $\widetilde{v} \in (V^*)^*$ given by $\widetilde{v}(\omega) = \omega(v)$ for $\omega \in V^*$. If $T\colon V \to V$ is any linear transformation, then there is an induced dual transformation $T^*\colon V^* \to V^*$, defined so that $T^*(\omega)(v) = \omega(T^{-1}v)$. In terms of the dual bases, if the linear transformation T has $n \times n$ matrix form A, then the dual map T^* has contragredient matrix form A^{-T}. Thus, given any representation ρ of a group G on V, we can construct the dual or contragredient representation ρ^* of G on V^*, where $\rho^*(g) = \rho(g)^*$ is obtained by applying the inverse transpose operation to the representation matrices.

Remark: In invariant theory, the vectors in $V = \mathbb{R}^n$ which transform according to the identity representation of $\mathrm{GL}(n)$ are known as *contravariant vectors*. In other words, a contravariant vector is one that is subject to our basic transformation rule $\overline{\mathbf{x}} = A\,\mathbf{x}$ for each $A \in \mathrm{GL}(n)$, cf. (2.4). Vectors in the dual space $V^* \simeq \mathbb{R}^n$, which transform according to the contragredient representation, are known (perhaps perversely, although we are adhering to the classical terminology) as *covariant vec-*

tors. The transformation rule for covariant vectors is $\mathbf{a} = A^T \bar{\mathbf{a}}$ or $\bar{\mathbf{a}} = A^{-T} \mathbf{a}$. For example, under contravariant transformations of its argument, the coefficients $\mathbf{a}$ of a linear form $Q(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} = \sum_i a_i x_i$ form the components of a covariant vector; see (4.5).

Next, recall that the tensor product $V \otimes W$ of two vector spaces can be identified with the space of linear maps $T: V^* \to W$ from the dual space of V to W. If $v \in V$ and $w \in W$, then there is an induced element $v \otimes w \in V \otimes W$, defined as the rank one linear transformation that takes $\omega \in V^*$ to the scalar multiple $\omega(v)\, w \in W$. If $\{e_1, \dots, e_n\}$ is a basis of V and $\{f_1, \dots, f_m\}$ a basis of W, then the tensor products $e_i \otimes f_j$ form a basis of $V \otimes W$, which has dimension mn. (In terms of the given bases and their duals, $e_i \otimes f_j$ corresponds to the linear transformation whose matrix form has 1 in the $(j,i)^{\text{th}}$ position.) In particular, $V \otimes \mathbb{R} \simeq V$, where we identify $v \otimes c$ with the scalar product cv. Given linear maps $T: V \to V$ and $U: W \to W$, we can define their tensor product $T \otimes U: V \otimes W \to V \otimes W$ in the obvious manner: $(T \otimes U)(v \otimes w) = T(v) \otimes U(w)$. In this way, the tensor product $\rho \otimes \sigma$ of representations ρ of G on V and σ of G on W defines a representation on the tensor product space $V \otimes W$. For example, the tensor product $\rho_e \otimes \rho_e$ of the standard representation of $\mathrm{GL}(n)$ with itself acts on the space $\mathbb{R}^n \otimes \mathbb{R}^n \simeq \mathcal{M}_{n \times n}$, which we identify with the space of $n \times n$ matrices. The reader can verify that this representation can be identified with the representation $X \mapsto AXA^T$ mentioned above. Similarly, the conjugation representation $X \mapsto AXA^{-1}$ is the same as the tensor product $\rho_e \otimes \rho_c$ of the standard representation with its dual.

Example 4.5. Consider the particular case when $G = \mathrm{GL}(2)$. An important example is provided by the tensor product of the contragredient representation

$$\rho_c \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \frac{1}{\alpha\delta - \beta\gamma} \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix}$$

with the determinantal representation $\rho_d \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \alpha\delta - \beta\gamma$. The representation space can be identified as $\mathbb{R}^2 \otimes \mathbb{R} \simeq \mathbb{R}^2$, and hence this tensor product defines the two-dimensional representation

$$(\rho_c \otimes \rho_d) \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \delta & -\gamma \\ -\beta & \alpha \end{pmatrix} = \tilde{A}.$$

However, the latter representation is, in fact, isomorphic to the identity representation $\rho_e(A) = A$. Indeed, introduction of the alternative basis

$\widetilde{e}_1 = e_2$, $\widetilde{e}_2 = -e_1$ produces

$$\widetilde{A}\,\widetilde{e}_1 = \widetilde{A}\,e_2 = -\beta e_1 + \alpha e_2 = \alpha\,\widetilde{e}_1 + \beta\,\widetilde{e}_2,$$
$$\widetilde{A}\,\widetilde{e}_2 = -\widetilde{A}\,e_1 = -\delta e_1 + \gamma e_2 = \gamma\,\widetilde{e}_1 + \delta\,\widetilde{e}_2,$$

and hence, in terms of the new basis, $\widetilde{A}$ has the same matrix form as A. In other words, in a two-dimensional vector space, contravariant vectors can be obtained from their covariant counterparts by multiplication by the determinantal representation.

Remark: This result is peculiar to GL(2) — there is no combination of contragredient and determinantal representations of GL(n) that reproduces the identity representation when $n > 2$. As a consequence, the invariant theory of GL(n) for $n > 2$ is rather more involved than the two-dimensional case. See Chapter 10 for further details.

Remark: The tensor product $V \otimes \mathbb{C}$ of a real vector space with the complex numbers defines a complex vector space known as the *complexification* of V. The complexification process, applied to real representations, proves to be a particularly powerful tool in the general theory.

Remark: The elements of the tensor powers of the vector space V are known as *contravariant tensors*, while the tensor powers of the dual V^* are known as *covariant tensors*. Binary forms can be regarded as symmetric covariant tensors. Elements of tensor products of one or more copies of the identity representation with one or more copies of the contragredient representation are known as *mixed tensors*. Example 4.5 implies that, in the two-dimensional situation, every type of tensorial representation can be obtained by tensoring a covariant tensor representation with by a suitable power of the determinantal representation, which amounts to an adjustment of the overall weight of the tensor.

Irreducibility

As we have just seen, many complicated group representations can be built up from simpler representations using the basic tensor operations. The simplest nontrivial cases are the irreducible representations, meaning those which admit no nontrivial subrepresentations.

Definition 4.6. Let ρ define a representation of a group G on a vector space V. A subspace $W \subset V$ is an *invariant subspace* if it has the property that $\rho(g)W \subset W$ for all $g \in G$.

Note that the restriction of ρ to an invariant subspace W induces a subrepresentation of G on W. Trivial invariant subspaces, valid for any representation, are $W = \{0\}$ and $W = V$.

Definition 4.7. A *reducible* representation is one that contains a nontrivial invariant subspace $\{0\} \neq W \subsetneq V$. An *irreducible* representation is one that has no nontrivial invariant subspaces.

Example 4.8. The direct sum $\rho \oplus \sigma$ of two representations is reducible since each appears as a subrepresentation therein.

Example 4.9. The identity representation ρ_e of $\mathrm{GL}(V)$ on V is clearly irreducible. However, none of its tensor powers, acting on the tensor product spaces $\bigotimes^k V = V \otimes \cdots \otimes V$, are irreducible. For example, if we identify $V \simeq \mathbb{R}^n$, then, as we remarked, the second tensor power $V \otimes V \simeq \mathcal{M}_{n\times n} \simeq \mathbb{R}^{n^2}$ can be identified with the space of $n \times n$ matrices, with representation $X \mapsto AXA^T$. Clearly the subspaces consisting of all symmetric matrices, $\bigodot^2 V = \{S^T = S\}$, and all skew-symmetric matrices, $\bigwedge^2 V = \{K^T = -K\}$, are invariant subspaces, and so $\bigotimes^2 V = \bigodot^2 V \oplus \bigwedge^2 V$ decomposes into a direct sum of symmetric and skew components, both of which form irreducible subrepresentations. Let us prove irreducibility in the symmetric case. Suppose $\{0\} \neq W \subset \bigodot^2 V$ is an invariant subspace, and let $0 \neq S \in W$ be a nonzero symmetric matrix therein. Using Sylvester's Theorem 3.56, we can diagonalize $ASA^T = S_0 = \mathrm{diag}(\pm 1, \dots, \pm 1, 0, \dots, 0) \in W$. Moreover, choosing $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ shows that $DS_0D^T = \mathrm{diag}(\pm\lambda_1^2, \dots, \pm\lambda_k^2, 0, \dots, 0)$ must lie in W. The permutation matrices constructed in Example 4.3 will act on such diagonal matrices by permuting the entries, so the nonzero entries can be placed anywhere on the diagonal. Since W is a subspace, we can take linear combinations of such diagonal matrices and arrive at the conclusion that W contains all diagonal matrices. But then Theorem 3.56 shows that every symmetric matrix is in W, proving the result. The skew-symmetric version is similar and left to the interested reader. (For instance, one can utilize the canonical forms for skew-symmetric matrices, cf. [**75**; §XI.4].)

The higher tensor powers of a vector space include the symmetric and skew-symmetric subspaces as irreducible subrepresentations, but these are not an exhaustive list. One indication of this fact is to note that $\bigotimes^3 V$ has dimension n^3, while the symmetric and completely skew-symmetric subspaces have respective dimensions $\frac{1}{6}n(n+1)(n+2)$ and $\frac{1}{6}n(n-1)(n-2)$, whose sum is less than n^3. The full classification of

the irreducible subrepresentations of the tensor representations of $\mathrm{GL}(V)$ can be found in [**231**], for instance, and lead to the theory of symmetry classes of tensors, classified by Young diagrams, which are also intimately related to the representation theory of the symmetric group $\mathbb{S}^n$. The resulting theory has important applications not only in mathematics, including combinatorics and geometry, cf. [**143, 231**], but also in quantum mechanics and chemistry, cf. [**232, 238**].

Exercise 4.10. Prove that if $\rho\colon G \to \mathrm{GL}(n)$ is any n-dimensional representation, then its determinant $\det\rho\colon G \to \mathbb{R} \setminus \{0\}$ defines a one-dimensional representation.

In many of the most important cases — for instance, when the group is finite, or compact, or the general tensor representations of $\mathrm{GL}(n)$ — finite-dimensional reducible representations can always be decomposed into a direct sum of irreducible subrepresentations, cf. [**231**]. For such groups, then, the irreducible representations form the fundamental building blocks, and it suffices to study their properties in order to understand completely general representations. The classification of irreducible representations has been the focus of major research efforts for over a half century; see [**144, 223, 231**] for example. Unfortunately, space precludes us from looking at anything beyond the particular examples that form the focus of this book.

Example 4.11. A simple counterexample to decomposability is provided by the reducible two-dimensional representation

$$\rho(A) = \begin{pmatrix} 1 & \log|\det A| \\ 0 & 1 \end{pmatrix}$$

of the general linear group $\mathrm{GL}(n, \mathbb{R})$. This representation leaves the subspace spanned by the vector $(1, 0)^T$ invariant, but there is no complementary invariant subspace, and hence ρ cannot be written as the direct sum of two subrepresentations.

Exercise 4.12. Prove that if a representation of G acts transitively on $V \setminus \{0\}$, then it is irreducible. Is the converse true?

Exercise 4.13. A representation is called *unitary*† if its image is contained in the group of norm-preserving linear transformations of an

† The term comes from the complex version, in which the representation maps G to the unitary group $\mathrm{U}(V)$ consisting of all norm-preserving linear transformations of the complex Hermitian inner product space V.

inner product space V, so that, in the real case, $\rho\colon G \to \mathrm{O}(V)$. Prove that any finite-dimensional unitary representation decomposes into a direct sum of irreducible subrepresentations. *Hint*: Prove that the orthogonal complement to an invariant subspace is also invariant.

The decomposability of the representations of finite groups is based on Exercise 4.13 combined with the following observation.

Proposition 4.14. *Every finite-dimensional representation of a finite group is equivalent to a unitary representation.*

Proof: Let us choose any inner product $\langle x\,;y\rangle$ on the representation space V. We then average over the group† to define a new inner product

$$\langle x\,;y\rangle_G = \frac{1}{\#G}\sum_{g\in G}\langle \rho(g)x\,;\rho(g)y\rangle \tag{4.2}$$

on V. It is easy to check that ρ is unitary with respect to the new group-invariant inner product. *Q.E.D.*

Remark: The same result holds for compact Lie groups (see Chapter 8), where one replaces the sum in (4.2) by an integral based on the invariant Haar measure, cf. [**212**, **231**]. Thus, Proposition 4.14 combined with Exercise 4.13 implies that any representation of a finite group or compact Lie group decomposes into a direct sum of irreducible subrepresentations.

Remark: A matrix group $G \subset \mathrm{GL}(n)$ is called *linearly reductive*, [**156**, **213**], if every representation that depends rationally on the matrix entries of $A \in G$ decomposes into a sum of irreducible representations. As we remarked, finite groups, compact Lie groups, and the groups $\mathrm{GL}(n)$ and $\mathrm{SL}(n)$ are all linearly reductive. The Hilbert Basis Theorem 2.42 holds for all linearly reductive groups, cf. [**156**], and hence one can argue that they form the optimal class of groups for generalizing the full range of methods and results in classical invariant theory.

Function Spaces

One easy way to turn a nonlinear group action into a linear representation is through its induced action on the functions defined on the space. Let G be a transformation group acting on a space X. Then there is a

† See (4.8) for a more detailed discussion of this process.

naturally induced representation of G on the linear space $\mathcal{F} = \mathcal{F}(X)$ of real-valued functions $F: X \to \mathbb{R}$. A group element $g \in G$ will map the function F to the transformed function $\overline{F} = g \cdot F$, which is defined by

$$\overline{F}(\bar{x}) = F(g^{-1} \cdot \bar{x}) = F(x). \tag{4.3}$$

The introduction of the inverse g^{-1} in (4.3) ensures that the action defines a group homomorphism: $g \cdot (h \cdot F) = (g \cdot h) \cdot F$ for all $g, h \in G$, and $F \in \mathcal{F}$. The representation of G on the (infinite dimensional) function space $\mathcal{F}$ will typically include a wide variety of important subrepresentations, including representations on subspaces of polynomials, continuous functions, smooth (differentiable) functions, analytic functions, normalizable (L^2) functions, and so on. This basic construction clearly extends to the space $\mathcal{F}^k(X)$ consisting of all vector-valued functions $F: X \to \mathbb{R}^k$.

For our invariant theoretic purposes, the most important examples are the standard representation of the general linear group $\mathrm{GL}(2, \mathbb{R})$ acting on $\mathbb{R}^2$, and its complex counterpart $\mathrm{GL}(2, \mathbb{C})$ acting on $\mathbb{C}^2$. Concentrating on the real version, the induced representation (4.3) on the space of real-valued functions $Q: \mathbb{R}^2 \to \mathbb{R}$ is explicitly given by

$$\overline{Q}(\bar{x}, \bar{y}) = \overline{Q}(\alpha\, x + \beta\, y, \gamma\, x + \delta\, y) = Q(x, y), \qquad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}(2). \tag{4.4}$$

In particular, if Q is a homogeneous polynomial, so is $\overline{Q}$, and hence the space $\mathcal{P}^{(n)} \subset \mathcal{F}(\mathbb{R}^2)$ of homogeneous polynomials of degree n forms an invariant subspace, and the representation (2.5) reduces to our original transformation rules (2.5) for binary forms. We let $\rho_n = \rho_{n,0}$ denote the induced finite-dimensional representation of $\mathrm{GL}(2)$ on $\mathcal{P}^{(n)}$. We can uniquely identify each homogeneous polynomial $Q \in \mathcal{P}^{(n)}$ with the $(n+1)$-tuple $\mathbf{a} = (a_0, a_1, \dots, a_n) \in \mathbb{R}^{n+1}$, where the a_i are its coefficients relative to the basis of $\mathcal{P}^{(n)} \simeq \mathbb{R}^{n+1}$ provided by the scaled monomials $\binom{n}{i} x^i y^{n-i}$, for $i = 0, \dots, n$. The explicit action of $\mathrm{GL}(2)$ on the coefficients a_i was given in (2.6). For example, the coefficients of a general linear polynomial $Q(x, y) = ax + by$ will transform according to

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \alpha & \gamma \\ \beta & \delta \end{pmatrix} \begin{pmatrix} \bar{a} \\ \bar{b} \end{pmatrix}, \qquad A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}(2), \tag{4.5}$$

and so form a covariant vector. Therefore, the representation ρ_1 on the space $\mathcal{P}^{(1)}$ can be identified with the contragredient representation of $\mathrm{GL}(2)$. On the space of quadratic forms $Q(x, y) = ax^2 + 2bxy + cy^2$, the induced representation ρ_2 acts on the coefficients as in (1.12),

which we identify with the second symmetric power of the contragredient representation, cf. Example 4.9.

A particularly important class of representation is obtained by tensoring the polynomial representations with powers of the determinantal representation. Under a group element, the function $\overline{Q}(\bar{x}, \bar{y})$ maps to $Q(x, y)$, where

$$\overline{Q}(\bar{x}, \bar{y}) = (\alpha\delta - \beta\gamma)^k \, Q(x, y) = (\alpha\delta - \beta\gamma)^k \, \overline{Q}(\alpha\, x + \beta\, y, \gamma\, x + \delta\, y). \tag{4.6}$$

The restriction of the representation (4.6) to the space $\mathcal{P}^{(n)}$ of homogeneous polynomials serves to define the fundamental representation $\rho_{n,k}$ of *weight* k and *degree* n, cf. (2.32). In particular, according to Example 4.5, the representation $\rho_{1,1}$ of weight 1 and degree 1 is isomorphic to the identity representation of GL(2).

Theorem 4.15. *The representation* $\rho_{n,k}$ *of* GL(2) *is irreducible.*

Proof: Let $\{0\} \neq W \subset \mathcal{P}^{(n)}$ be a nonzero invariant subspace. We first show that if $Q(x, y)$ is any polynomial in W, as in (2.1), then every monomial appearing in Q with nonzero coefficient also lies in W; that is, $x^i y^{n-i} \in W$ provided $a_i \neq 0$. This follows immediately from the invariance of W under the unimodular scaling $(x, y) \mapsto (\lambda x, \lambda^{-1} y)$, which changes Q into

$$\widehat{Q}_\lambda(x, y) = \sum_{i=0}^{n} \binom{n}{i} a_i \lambda^{2i-n} \, x^i y^{n-i} = \lambda^{-n} \sum_{i=0}^{n} \binom{n}{i} a_i \lambda^{2i} \, x^i y^{n-i}. \tag{4.7}$$

Now $\widehat{Q}_\lambda \in W$ for each $\lambda \in \mathbb{R}$. But $\lambda^n \, \widehat{Q}_\lambda$ is a polynomial in the parameter λ, hence this will hold if and only if each coefficient of each power of λ itself lies in W, proving the assertion.

Thus, we have proved that any subspace invariant under the unimodular scaling subgroup must be spanned by monomials. Next, suppose $x^k y^{n-k} \in W$ is any monomial. We apply the linear transformation $(x, y) \mapsto (x + \beta y, y)$, which maps it to $\sum_{i=0}^{k} \binom{k}{i} \beta^i \, x^i y^{n-i}$, which, as before, lies in W, and hence each of its constituent monomials, $x^k y^{n-k}$, $x^{k-1} y^{n-k+1}, \dots, y^n$, also lies in W. Application of the linear transformation $(x, y) \mapsto (x, \gamma x + y)$ shows that the monomials $x^n, \dots, x^k y^{n-k}$ also lie in W. Therefore W contains every monomial of degree n, and hence $W = \mathcal{P}^{(n)}$, which proves irreducibility. *Q.E.D.*

Remark: In fact, the representations $\rho_{n,k}$ provide a complete list of all irreducible finite-dimensional representations of GL(2). Moreover,

any tensor representation can be decomposed into a direct sum of copies of the $\rho_{n,k}$; see [**231**] for a proof.

Note that our proof utilized only unimodular linear transformations, proving that $\rho_{n,k}$ remains irreducible when restricted to the special linear group $\mathrm{SL}(2)$. On the other hand, if $n \geq 2$, the restriction of the representation $\rho_{n,k}$ to the rotation subgroup $\mathrm{SO}(2) \subset \mathrm{GL}(2)$ is no longer irreducible. For example, the representation of $\mathrm{SO}(2)$ on the space of real quadratic polynomials $\mathcal{P}^{(2)}$ decomposes into the sum of two irreducible subrepresentations, a trivial one on the one-dimensional subspace spanned by $x^2 + y^2$, and a two-dimensional representation on the subspace spanned by $x^2 - y^2$ and xy. Finally, we remark that the preceding proof shows that the subgroup of upper triangular unimodular matrices is *not* linearly reductive since each of the subspaces $W_k = \mathrm{Span}\{x^k y^{n-k}, x^{k-1}y^{n-k+1}, \dots, y^n\}$ is invariant, but clearly not irreducible (unless $k = 0$).

Exercise 4.16. Decompose the restriction of $\rho_{n,k}$ to the subgroup $\mathrm{SO}(2) \subset \mathrm{GL}(2)$ into irreducible subrepresentations. *Hint*: Note that the subspaces $\mathcal{H}_m = \{(x^2+y^2)^m Q(x,y) \,|\, Q \in \mathcal{P}^{(n-2m)}\}$ are rotationally invariant.

Exercise 4.17. The (complex) one-dimensional representations of the circle group $\mathrm{SO}(2) \simeq S^1$ are given by $\rho_n(\theta) = e^{in\theta}$, where θ is the angular coordinate on S^1 and $n \in \mathbb{Z}$. It can be proved that any other (complex) representation can be decomposed into a direct sum of these irreducible representations. For example, prove that the standard representation $\rho(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ decomposes into the direct sum $\rho_1 \oplus \rho_{-1}$. Similarly, show that the representation $\rho^{(2)}$ of $\mathrm{SO}(2)$ on the space of complex quadratic polynomials breaks up into the direct sum $\rho_2 \sim \rho_0 \oplus \rho_{-2}$. *Note*: The subspace $W = \mathrm{Span}\{x^2 - y^2, xy\} \subset \mathcal{P}^{(2)}$ mentioned above is irreducible over the reals but decomposes into two invariant one-dimensional subspaces over the complexes.

Remark: The appearance of the Fourier modes $e^{in\theta}$ in this case is not an accident, but forms a special case of a general theory, due to Peter and Weyl, [**179**], of Fourier analysis on compact Lie groups. Details, including applications to the theory of special functions, can be found in [**212**].

Invariant Functions

In general, an invariant is defined as a real-valued function that is unaffected by group transformations. The determination of a complete set of invariants of a given group action is a problem of supreme importance for the study of equivalence and canonical forms. For example, in sufficiently regular cases, the orbits, and hence the canonical forms, for a group action are completely characterized by its invariants; see Theorem 8.17.

Definition 4.18. Let G be a transformation group acting on a space X. An *invariant* is a real-valued function $I: X \to \mathbb{R}$ which satisfies $I(g \cdot x) = I(x)$ for all transformations $g \in G$.

In other words, an invariant function is a fixed point for the induced representation of G on the function space $\mathcal{F}(X)$.

Proposition 4.19. *Let $I: X \to \mathbb{R}$. The following conditions are equivalent:*

(a) *I is a G-invariant function.*

(b) *I is constant on the orbits of G.*

(c) *All level sets $\{I(x) = c\}$ are G-invariant subsets of X.*

In particular, constant functions are trivially G-invariant. If G acts transitively on X, then these are the only invariants.

Example 4.20. For the standard action (3.11) of the rotation group $\mathrm{SO}(2)$ on $\mathbb{R}^2$, the radius $r = \sqrt{x^2 + y^2}$ is an invariant function, as is any function $f(r)$ thereof. Indeed, Proposition 4.19 immediately implies that these are the only invariants for the rotation group. Note that in this case the invariant function r serves to distinguish the orbits: two points lie in the same circular orbit if and only if they have the same value for the radial invariant.

Example 4.21. Even though there are several orbits, there are no continuous, nonconstant invariant functions for the action of $\mathrm{GL}(2)$ on the space $\mathcal{Q}$ of real quadratic forms described in Example 3.53, and hence the orbits are *not* distinguished by continuous invariants. However, we can certainly find discontinuous invariant functions, e.g. the sign of the discriminant,† which will serve the purpose.

† The discriminant itself is an $\mathrm{SL}(2)$ invariant, but is not $\mathrm{GL}(2)$–invariant.

A fundamental problem is to determine *all* possible (continuous, polynomial, analytic, etc.) invariants of a group of transformations. Note that if $I_1(x), \dots, I_k(x)$ are invariants, and $H(y_1, \dots, y_k)$ is any function, then $I(x) = H(I_1(x), \dots, I_k(x))$ is also invariant. Therefore, we only need to find a complete set of functionally independent invariants, having the property that any other invariant can be written as a function thereof. In the case of polynomials, one can ask for more detailed information: algebraically independent invariants, as in the definition of a Hilbert basis, rationally independent invariants, and so on.

For finite groups, one can introduce a method of "group averaging" to generate invariants. Let G be a finite group with $\#G$ elements that acts on a space X. The *symmetrization* or *Reynolds operator*

$$\varsigma = \frac{1}{\#G} \sum_{g \in G} g \tag{4.8}$$

averages functions over the group and so defines a projection from the space $\mathcal{F}(X)$ of all functions on X to the subspace of G-invariant functions. In other words, if $F\colon X \to \mathbb{R}$ is any function, then $I(x) = \varsigma \cdot F(x) = (\#G)^{-1} \sum F(g \cdot x)$ is an invariant; moreover, if I is already invariant, then $\varsigma \cdot I = I$.

Example 4.22. Consider the representation of the symmetric group $\mathbb{S}^n$ on $\mathbb{R}^n$ given by the permutation matrices, as in Example 4.3. A function $I\colon \mathbb{R}^n \to \mathbb{R}$ is invariant under $\mathbb{S}^n$ if and only if it satisfies

$$I(x_{\pi(1)}, \dots, x_{\pi(n)}) = I(x_1, \dots, x_n), \qquad \text{for every} \qquad \pi \in \mathbb{S}^n. \tag{4.9}$$

The invariants of the permutation group $\mathbb{S}^n$ are known as *symmetric functions*. According to (4.8), the associated symmetrization operator

$$\varsigma = \frac{1}{n!} \sum_{\pi \in \mathbb{S}^n} \pi \tag{4.10}$$

defines the projection onto the subspace of all symmetric functions. Of particular importance are the elementary symmetric polynomials

$$\begin{aligned} \sigma_1(x_1, \dots, x_n) &= x_1 + \cdots + x_n, \\ \sigma_2(x_1, \dots, x_n) &= x_1x_2 + x_1x_3 + \cdots + x_{n-1}x_n = \sum_{i<j} x_i x_j, \\ &\vdots \\ \sigma_n(x_1, \dots, x_n) &= x_1x_2 \cdots x_n. \end{aligned}$$

In general,

$$\sigma_k(x_1, \dots, x_n) = \binom{n}{k} \varsigma(x_1 x_2 \cdots x_k) = \sum_{i_1 < i_2 < \cdots < i_k} x_{i_1} x_{i_2} \cdots x_{i_k}. \quad (4.11)$$

The Fundamental Theorem for the symmetric group states that the elementary symmetric polynomials form a complete system of invariants.

Theorem 4.23. *If $I(x)$ is a symmetric polynomial, then I can be written uniquely in terms of the elementary symmetric polynomials*: $I(x) = P(\sigma_1(x), \cdots, \sigma_n(x))$.

Proof: Let us order the monomials $x^K = x_1^{k_1} \cdots x_n^{k_n}$ *lexicographically*, so that $x^K < x^J$ whenever the first nonvanishing difference $j_1 - k_1, j_2 - k_2, \dots$ is positive. The last monomial in the elementary symmetric polynomial σ_k in the lexicographic ordering is $x_1 x_2 \cdots x_k$. Given a symmetric polynomial $I(x)$, let $a_K x^K$, $a_K \neq 0$, be its last monomial. Since $I(x)$ is symmetric, the permuted monomials $x^{\pi(K)}$ also occur with the same nonvanishing coefficient a_K; therefore, $x^{\pi(K)} < x^K$, which implies that $k_1 \geq k_2 \geq \cdots \geq k_n$. On the other hand, the last monomial in the power product of elementary symmetric polynomials $S^{\overline{K}} = \sigma_1^{k_1 - k_2} \sigma_2^{k_2 - k_3} \cdots \sigma_{n-1}^{k_{n-1} - k_n} \sigma_n^{k_n}$ is also x^K. The difference $\widetilde{I} = I - a_K S^{\overline{K}}$ is also symmetric and has a last monomial of lower order. A straightforward induction completes the demonstration that I can be written as a polynomial in the elementary symmetric polynomials. In order to prove uniqueness, it suffices to note that each power product $S^{\overline{K}}$ of elementary symmetric polynomials has a different last monomial, namely x^K. *Q.E.D.*

Exercise 4.24. The *power sum symmetric polynomials* are obtained by symmetrizing $(x_1)^k$, so $P_k(x) = \varsigma((x_1)^k) = \sum (x_i)^k$. Use the method of proof of Theorem 4.23 to express the P_k in terms of the elementary symmetric polynomials for k small. Can you extend your result to general k? See [**143**] for details.

Exercise 4.25. Prove that the coefficients c_k in a polynomial (2.16) of degree n can be written in terms of the elementary symmetric polynomials of its (complex) roots p_k, as in (2.17). Specifically,

$$c_k = (-1)^{n-k} c_n \sigma_k(p_1, \dots, p_n), \qquad k = 1, \dots, n.$$

Remark: Extensions of the Fundamental Theorem 4.23 to more general symmetric functions are discussed in [**14, 16, 81**].

Joint Invariants

Joint invariants appear when a transformation group acts simultaneously on several different spaces or, more typically, on multiple copies of the same space. More specifically, suppose G is a fixed group which acts on the spaces $X_1, \dots, X_m$. Then there is a naturally induced action of G on the Cartesian product space $X_1 \times \cdots \times X_m$ given by $g \cdot (x_1, \dots, x_m) = (g \cdot x_1, \dots, g \cdot x_m)$ for $x_i \in X_i$, $g \in G$.

Definition 4.26. A *joint invariant* is merely an invariant function $J: X_1 \times \cdots \times X_m \to \mathbb{R}$ for a Cartesian product action of a group G. In other words, $J(g \cdot x_1, \dots, g \cdot x_m) = J(x_1, \dots, x_m)$ for all $g \in G$, $x_i \in X_i$.

When the spaces $X_i = X$ are identical, with identical actions of G, we shall call a joint invariant $J(x_1, \dots, x_m)$ depending on m points $x_i \in X$ an *m-fold joint invariant*. Any ordinary invariant $I(x)$ induces trivial m–fold joint invariants, namely, $J_i(x_1, \dots, x_m) = I(x_i)$ for any $1 \leq i \leq m$. Typically, one is really only interested in "genuine" joint invariants, which are not trivially obtained from joint invariants depending on fewer than m points. Vice versa, if $J(x_1, \dots, x_m)$ is an m–fold joint invariant, then its *trace*

$$I(x) = J(x, x, \dots, x), \qquad x \in X, \tag{4.12}$$

produces an ordinary invariant on X.

Example 4.27. Consider the Euclidean group $\mathrm{E}(2)$ acting on the plane $X = \mathbb{R}^2$. Since the action is transitive, there are no ordinary invariants. On the Cartesian product $X^{(2)} = X \times X$, there is a single joint invariant, namely, the distance $d(\mathbf{x}_1, \mathbf{x}_2) = \| \mathbf{x}_1 - \mathbf{x}_2 \|$ between two points $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$. On higher Cartesian powers, $X^{(m)} = X \times \cdots \times X$, the two-fold joint invariant induces a variety of joint invariants:

$$d_{ij} = d(\mathbf{x}_i, \mathbf{x}_j) = \| \mathbf{x}_i - \mathbf{x}_j \| \tag{4.13}$$

for any $i < j$. A fundamental result states that these are the only joint invariants — any other joint invariant $I(\mathbf{x}_1, \dots, \mathbf{x}_m)$ can be written as a function of the interpoint distances. See Example 8.29, and also Weyl, [**231**; Theorem 2.9.A], for more details.

Example 4.28. A similar result holds for the equi-affine group, $\mathrm{SA}(2)$, consisting of area-preserving affine transformations, as introduced in Example 3.36. In this case there are no (continuous) ordinary

or two-fold joint invariants, so the first example of a joint invariant arises on $X^{(3)}$, and is given by a_{123}, where†

$$a_{ijk} = A(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) = \tfrac{1}{2}(\mathbf{x}_j - \mathbf{x}_i) \wedge (\mathbf{x}_j - \mathbf{x}_k) \tag{4.14}$$

denotes the (signed) area of the triangle with corners $\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k$. Again, in Example 8.31 we prove that every other joint invariant can be written as a function of the three-fold area invariants (4.14). Note that the area a_{ijk} is only a proper Euclidean joint invariant, since a reflection will reverse its sign; however, its square is Euclidean–invariant. According to the preceding example, then, we should be able to write the latter in terms of the distances between the vertices of the triangle. Of course, the answer is provided by the well-known semi-perimeter formula $(a_{ijk})^2 = s(s - d_{ij})(s - d_{ik})(s - d_{jk})$, where $s = \frac{1}{2}(d_{ij} + d_{ik} + d_{jk})$ is the semi-perimeter of the triangle with vertices $\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k$.

For the full affine group A(2), the area of a triangle is multiplied by the determinant of the linear part of the transformation and hence forms a "relative joint invariant" — see below. The ratio $r_{ijkl} = a_{ijk}/a_{ijl}$ between two such areas provides a genuine joint affine invariant. Again, all joint affine invariants are expressible as functions of these ratios.

Exercise 4.29. Determine the orbits of the action of the groups SA(2) and A(2) acting on $\mathbb{R}^2 \times \mathbb{R}^2$ and $\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2$.

Example 4.30. The next example is of crucial importance for understanding how the group of linear transformations affects the geometry of the roots of polynomials. Consider the joint projective action of $\mathrm{SL}(2, \mathbb{C})$ on the m-fold Cartesian product $\mathbb{CP}^1 \times \cdots \times \mathbb{CP}^1$ given by

$$(p_1, \dots, p_m) \longmapsto \left(\frac{\alpha p_1 + \beta}{\gamma p_1 + \delta}, \dots, \frac{\alpha p_m + \beta}{\gamma p_m + \delta} \right), \quad \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}(2, \mathbb{C}).$$

Let us concentrate on the open subset $\widetilde{X}^{(m)} = \{p_i \neq p_j, i \neq j\}$ consisting of distinct m-tuples of points. For $m \leq 3$, the action is transitive on $\widetilde{X}^{(m)}$, and there are no nonconstant two- or three-fold invariants. Indeed, we can map any three distinct points (p_1, p_2, p_3) on the Riemann sphere to any desired canonical form, e.g., $(0, 1, \infty)$, by a suitable linear fractional transformation. The *cross-ratio*

$$[\, p_1, p_2, p_3, p_4 \,] = \frac{(p_1 - p_2)(p_3 - p_4)}{(p_1 - p_3)(p_2 - p_4)} \tag{4.15}$$

† The symbol $\wedge$ denotes the standard scalar cross product between planar vectors: $(x, y) \wedge (x', y') = xy' - x'y$.

is four-fold joint invariant, as can be verified directly. (If one of the points is infinite, (4.15) is computed in a consistent manner.) Now, only the first three points can be fixed, so a canonical form for four points could be $(0, 1, \infty, z)$, where the value of the projective modulus z is fixed by the cross-ratio $[0, 1, \infty, z] = 1/(1-z)$. In Example 8.34 we shall prove that every joint projective invariant can be written as a function of the cross-ratios of the points taken four at a time.

Exercise 4.31. Discuss the action of $\mathrm{SL}(2, \mathbb{R})$ on $\mathbb{RP}^1 \times \cdots \times \mathbb{RP}^1$, and on $\mathbb{CP}^1 \times \cdots \times \mathbb{CP}^1$.

Example 4.32. A joint invariant for the standard action of the symmetric group $\mathbb{S}^n$ on vectors in $\mathbb{R}^n$, as in Example 4.3, is known as a *multi-symmetric function*. The theory of multi-symmetric polynomials is less extensively investigated than the more elementary theory of symmetric polynomials, although it has been the subject of intermittent research activity, beginning with Junker, [**121**, **122**], and Netto, [**158**; §377–386]. See [**147**, **80**], for connections with combinatorics, [**5**] for recent applications to algebraic topology, and [**172**] for applications to dissipative decompositions of partial differential equations.

We consider the Cartesian product representation of $\mathbb{S}^n$ on the space $\mathbb{R}^n \times \cdots \times \mathbb{R}^n \simeq \mathbb{R}^{mn}$ consisting of m copies of $\mathbb{R}^n$. A function $Q(\mathbf{x}^1, \dots, \mathbf{x}^m)$, where $\mathbf{x}^i = (x^i_1, \dots, x^i_n) \in \mathbb{R}^n$, is called a multi-symmetric function if $\pi \cdot Q = Q$ for every $\pi \in \mathbb{S}^n$, which acts by simultaneous permutation of the components of the vectors $\mathbf{x}_i$. Alternatively, one can view a multi-symmetric function $\widetilde{Q}(\widetilde{\mathbf{x}}_1, \dots, \widetilde{\mathbf{x}}_n)$ as depending on n vectors $\widetilde{\mathbf{x}}_k = (x^1_k, \dots, x^m_k) \in \mathbb{R}^m$ and require symmetry under permutations of the vectors: $\widetilde{Q}(\widetilde{\mathbf{x}}_{\pi(1)}, \dots, \widetilde{\mathbf{x}}_{\pi(n)}) = \widetilde{Q}(\widetilde{\mathbf{x}}_1, \dots, \widetilde{\mathbf{x}}_n)$.

We can evidently construct multi-symmetric functions by applying the symmetrization operator (4.10) to ordinary functions. In particular, the *elementary multi-symmetric polynomials* are found by applying ς to the multi-linear monomials:

$$\sigma_I(\mathbf{x}^1, \dots, \mathbf{x}^m) = \binom{n}{k} \varsigma(x^{i_1}_1 x^{i_2}_2 \cdots x^{i_k}_k), \qquad \begin{matrix} I = (i_1, \dots, i_k), \\ 1 \le i_\kappa \le n. \end{matrix} \tag{4.16}$$

For example, in the case $m = n = 2$, the multi-symmetric functions depend on two sets of two variables and satisfy the symmetry condition $Q(x_2, x_1; y_2, y_1) = Q(x_1, x_2; y_1, y_2)$. The linear multi-symmetric functions are

$$\sigma_1 = x_1 + x_2, \qquad \sigma_2 = y_1 + y_2.$$

There are three quadratic elementary multi-symmetric functions:

$$\sigma_{11} = x_1 x_2, \qquad \sigma_{12} = \tfrac{1}{2}(x_1 y_2 + x_2 y_1), \qquad \sigma_{22} = y_1 y_2.$$

As before, the Fundamental Theorem states that the elementary multi-symmetric polynomials form a generating set.

Theorem 4.33. *Any multi-symmetric polynomial can be written as a polynomial in the elementary multi-symmetric polynomials.*

The proof of this result is similar to that of Theorem 4.23; see [**80**, **158**]. There is, however, one crucial difference between the two versions — the formula expressing a multi-symmetric polynomial in terms of the elementary ones is *not* unique, owing to the existence of nontrivial syzygies. The simplest such syzygy is in the case $n = 2$, where

$$4(\sigma_{ij}\sigma_{kl} - \sigma_{ik}\sigma_{jl}) = \sigma_i\sigma_j\sigma_{kl} + \sigma_k\sigma_l\sigma_{ij} - \sigma_i\sigma_k\sigma_{jl} - \sigma_j\sigma_l\sigma_{ik}. \tag{4.17}$$

Thus, in the case $m = n = 2$, there is one such syzygy,

$$4\big[\sigma_{11}\sigma_{22} - (\sigma_{12})^2\big] = (\sigma_2)^2\sigma_{11} - 2\sigma_1\sigma_2\sigma_{12} + (\sigma_1)^2\sigma_{22}. \tag{4.18}$$

It can be shown, [**63**], that (4.17) forms a complete list of syzygies when $n = 2$. Higher order syzygies are more complicated, and still not completely classified,† although particular cases appear in [**121**, **122**, **172**].

Multiplier Representations

Although induced actions of transformation groups on functions provide us with a wide variety of important representations, these are not quite general enough for our purposes. Consider the linear fractional action (2.7) of the group $\mathrm{GL}(2, \mathbb{C})$ on the projective line $\mathbb{CP}^1$. According to the preceding construction, (4.3), this induces the representation

$$\overline{F}(\bar{p}) = \overline{F}\left(\frac{\alpha p + \beta}{\gamma p + \delta}\right) = F(p)$$

† A personal story: In the first version of [**172**], unaware of the work of Junker, Cheri Shakiban and I ran several computer algebra computations of multi-symmetric syzygies. Using the specialized MACAULAY computer algebra package on a workstation, we were just able to treat the cases when $m, n \leq 3$, but this came close to the limits of the machine at that time. However, the referee of our paper pointed out Junker's thesis (written under Hilbert), which successfully treated cases like $m = 2, n = 7$, and $m = 3, n = 4$. So much for the power of symbolic computing!

on the space of scalar-valued functions on $\mathbb{CP}^1$, written in terms of the projective coordinate $p = x/y$. However, with the exception of the constants, which are homogeneous of degree 0, this representation does not naturally contain any of the fundamental polynomial representations $\rho_{n,k}$ given by (4.6). Indeed, each homogeneous polynomial (2.1) has an inhomogeneous representative given by (2.2). The induced action of GL(2) corresponding to the representation $\rho_{n,k}$ is readily seen to be

$$\begin{aligned} Q(p) &= (\alpha\delta - \beta\gamma)^k \, (\gamma p + \delta)^n \, \overline{Q}(\bar{p}) \\ &= (\alpha\delta - \beta\gamma)^k \, (\gamma p + \delta)^n \, \overline{Q}\left(\frac{\alpha p + \beta}{\gamma p + \delta}\right). \end{aligned} \tag{4.19}$$

The linear action (4.19) defines a more general kind of representation of the projective group on the space of functions over $\mathbb{CP}^1$, which is known as a "multiplier representation", and the prefactor $(\alpha\delta - \beta\gamma)^k (\gamma p + \delta)^n$, or, rather, its reciprocal, is called the *multiplier*. The general definition† follows, cf. [**17**, **153**, **231**].

Definition 4.34. Let G be a group acting on a space X. A *multiplier representation* of G is a representation $\overline{F} = g \cdot F$ on the space of real-valued functions $\mathcal{F}(X)$ of the particular form

$$\overline{F}(\bar{x}) = \overline{F}(g \cdot x) = \mu(g, x) \, F(x), \qquad g \in G, \quad F \in \mathcal{F}. \tag{4.20}$$

The condition that (4.20) actually defines a representation of the group G requires that the multiplier μ satisfy a certain algebraic identity, which follows directly from the group law $g \cdot (h \cdot F) = (g \cdot h) \cdot F$.

Lemma 4.35. *A function $\mu \colon G \times X \to \mathbb{C} \setminus \{0\}$ is a multiplier for a transformation group G acting on a space X if and only if it satisfies the multiplier equation*

$$\begin{aligned} &\mu(g \cdot h, x) = \mu(g, h \cdot x) \, \mu(h, x), \\ &\mu(e, x) = 1, \end{aligned} \qquad \text{for all} \quad \begin{cases} g, h \in G, \\ x \in X. \end{cases} \tag{4.21}$$

Note that if $\mu(g, x)$ and $\widehat{\mu}(g, x)$ are multipliers for a group G, so is their product $\mu(g, x) \cdot \widehat{\mu}(g, x)$, as is any power $\mu(g, x)^k$.

Remark: If a multiplier does not depend on the point x, so $\mu = \mu(g)$, then the multiplier equation (4.21) reduces to

$$\mu(g \cdot h) = \mu(g) \, \mu(h), \quad \text{for all} \quad g, h \in G, \qquad \text{and} \qquad \mu(e) = 1.$$

---

† This definition of multiplier representation is *not* the same as that appearing in the work of Mackey, [**144**]; the latter are also known as projective representations, [**100**], and will not play any role in our discussion.

In other words, μ defines a one-dimensional representation of the group G. Weyl, [**231**], only permits such multipliers, but (4.19) underscores the need to allow x dependence in our treatment.

Example 4.36. Consider the usual projective action (2.7) of the general linear group GL(2). By the preceding remark, the determinantal representation $\mu_{0,1}(A) = \alpha\delta - \beta\gamma$, for $A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, automatically defines a multiplier. More interestingly, the function

$$\mu_1\left(\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, p\right) = \gamma p + \delta$$

also defines a multiplier.† Indeed, $\mu_1(\mathbb{1}, p) = 1$, and

$$\begin{aligned} \mu_1\left(\begin{pmatrix} \alpha\tilde{\alpha} + \beta\tilde{\gamma} & \alpha\tilde{\beta} + \beta\tilde{\delta} \\ \gamma\tilde{\alpha} + \delta\tilde{\gamma} & \gamma\tilde{\beta} + \delta\tilde{\delta} \end{pmatrix}, p\right) &= (\gamma\tilde{\alpha} + \delta\tilde{\gamma})p + (\gamma\tilde{\beta} + \delta\tilde{\delta}) \\ = \gamma(\tilde{\alpha}p + \tilde{\beta}) + \delta(\tilde{\gamma}p + \tilde{\delta}) &= \left[\gamma\left(\frac{\tilde{\alpha}p + \tilde{\beta}}{\tilde{\gamma}p + \tilde{\delta}}\right) + \delta\right](\tilde{\gamma}p + \tilde{\delta}) \\ = \mu_1\left(\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, \frac{\tilde{\alpha}p + \tilde{\beta}}{\tilde{\gamma}p + \tilde{\delta}}\right) &\cdot \mu_1\left(\begin{pmatrix} \tilde{\alpha} & \tilde{\beta} \\ \tilde{\gamma} & \tilde{\delta} \end{pmatrix}, p\right), \end{aligned}$$

as required by (4.21). Since we can multiply multipliers, the functions

$$\mu_{n,k}\left(\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}, p\right) = (\alpha\delta - \beta\gamma)^{-k}(\gamma p + \delta)^{-n} \tag{4.22}$$

also define multipliers for the projective group. The restriction of this multiplier representation to the space $\mathcal{P}^n$ of polynomials of degree $\leq n$ coincides with the fundamental representation $\rho_{n,k}$ defined by (4.19).

There is a trivial way to obtain multipliers from the ordinary representation (4.3) of a transformation group G on the function space $\mathcal{F}(X)$. Suppose we multiply every function by a fixed nonvanishing function $\eta\colon X \to \mathbb{C} \setminus \{0\}$, and set $F^*(x) = \eta(x)\,F(x)$ for each $F \in \mathcal{F}$. The function η is sometimes known as a *gauge factor*.‡ In terms of the

† Note that μ_1 vanishes at the point $p_0 = -\gamma/\delta$, but the projective group transformation maps p_0 to ∞, so consistency is maintained.

‡ This terminology comes from physics and has its origins in Weyl's speculative attempt, [**233**], to unify electromagnetism and gravity. Our use of the term "gauge" is closer in spirit to Weyl's original coinage since, unlike the modern definition appearing in quantum electrodynamics, cf. [**232, 22**], our gauge factors $\eta(x)$ are not restricted to be of modulus 1. See [**85, 169**] for further developments and applications.

new choice of gauge, then, the standard representation (4.3) takes the modified form

$$\overline{F}^*(\bar{x}) = \frac{\eta(\bar{x})}{\eta(x)}\, F^*(x), \qquad \text{where} \qquad \bar{x} = g\cdot x, \quad \overline{F}^* = g\cdot F^*.$$

The function $\mu(g,x) = \eta(g\cdot x)/\eta(x)$ is readily seen to satisfy the multiplier equation (4.21) and so defines a *trivial multiplier*.

Definition 4.37. Two multipliers $\mu, \widetilde{\mu}\colon G\times X \to \mathbb{C}$ are *gauge equivalent* if they are related by the formula

$$\widetilde{\mu}(g,x) = \frac{\eta(g\cdot x)}{\eta(x)}\,\mu(g,x), \tag{4.23}$$

for some nonzero function $\eta\colon X \to \mathbb{C}\setminus\{0\}$.

Equivalent multipliers prescribe the same underlying multiplier representation, up to multiplication by a function. For example, it can be shown, cf. [**169**], that the multipliers (4.22) provide a complete list of inequivalent multipliers for the projective action of the group GL(2).

Exercise 4.38. Consider the standard linear action of the group SL(2) on $\mathbb{R}^2$ given by matrix multiplication. Prove that, for $k\in\mathbb{R}$,

$$\mu_k\left(\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}; x, y\right) = \exp\left[\frac{k\,\gamma}{y(\gamma x+\delta y)}\right], \qquad \alpha\delta-\beta\gamma = 1, \tag{4.24}$$

defines a multiplier. In fact, these give a complete list of inequivalent multipliers for this group action, cf. [**84**], [**169**; Exercise 3.25]. Does the multiplier (4.24) extend to the full general linear group GL(2)?

Although we shall primarily deal with scalar multipliers, the definition readily extends to vector-valued functions. On the function space $\mathcal{F}^k(X)$, a multiplier will be a matrix-valued function $\mu\colon G\times X\to \mathrm{GL}(k)$ that satisfies the multiplier equation (4.21); the order of the factors is now important in the first condition, and we replace 1 by the $k\times k$ identity matrix $\mathbb{1}$ in the second. In particular, any representation $\rho\colon G\to \mathrm{GL}(k)$ defines a matrix multiplier $\mu = \rho(g)$ that does not depend on the spatial coordinates x.

Exercise 4.39. Suppose G acts smoothly on $X\subset\mathbb{R}^n$, with explicit transformation formulae $g\cdot x = w(g,x)$. Prove that the Jacobian matrix $J(g,x) = \partial w(g,x)/\partial x$ defines a matrix-valued multiplier. Use this to conclude that the Jacobian determinant $\mu(g,x) = \det J(g,x)$ forms a scalar multiplier for any transformation group. Determine the Jacobian multiplier for the linear fractional action of GL(2).

Relative Invariants

Earlier we characterized an invariant of a group of transformations as a fixed point of the induced representation on the space of functions: $g \cdot I = I$ for all $g \in G$. The analog of an invariant for a multiplier representation is known as a relative invariant.

Definition 4.40. Let $\mu: G \times X \to \mathbb{R}$ be a multiplier for a transformation group action. A *relative invariant* of *weight* μ is a function $R: X \to \mathbb{R}$ which satisfies

$$R(g \cdot x) = \mu(g, x)\, R(x). \tag{4.25}$$

It is not hard to see that, as long as $R \not\equiv 0$ satisfies (4.25), the weight function μ must necessarily be a multiplier for the group action. For clarity, ordinary invariants are, occasionally, referred to as *absolute invariants*. For example, the invariants appearing in classical invariant theory are absolute invariants for the unimodular group SL(2), but only relative invariants for the full group GL(2). The relevant multiplier is the determinantal power $\mu_k(A) = (\det A)^{-k}$, where k determines the weight of the invariant. Thus, (2.19) is a particular case of the general condition (4.25) for a relative invariant.

If R and S are relative invariants corresponding to the *same* multiplier μ, then any linear combination $c_1 R + c_2 S$ is also a relative invariant of weight μ. (However, this certainly does not hold if R and S are relative invariants corresponding to different multipliers!) If R has weight μ and S has weight ν, then their product $R \cdot S$ is a relative invariant for the product multiplier $\mu \cdot \nu$. In particular, the ratio R/S of two relative invariants having the *same* weight is an absolute invariant of the group. Therefore, once we know one relative invariant of a given weight μ, we can easily provide a complete list of all such relative invariants.

Proposition 4.41. *Let μ be a scalar multiplier for a transformation group G. If $R_0(x)$ is a nonvanishing relative invariant of weight μ, then every other relative invariant of weight μ has the form $R(x) = I(x) R_0(x)$, where I is any absolute invariant.*

Remark: Proposition 4.41 does *not* guarantee the existence of a nontrivial relative invariant. In the case of matrix multipliers, one typically requires several different relative invariants to form a basis (with respect to the absolute invariants) for the space of vector-valued relative invariants. See [**68**] for a general theorem that characterizes the precise number of relative invariants for "regular" matrix multipliers.

The classical theory can also be fitted into this general framework. A classical invariant $I(\mathbf{a})$ is a relatively invariant function $I\colon \mathcal{P}^{(n)} \to \mathbb{C}$ defined on the space of homogeneous polynomials of degree n under the induced representation of GL(2). A covariant $J(\mathbf{a}, \mathbf{x})$ can be viewed as a homogeneous function $J\colon \mathcal{P}^{(n)} \times \mathbb{C}^2 \to \mathbb{C}$ which is a *joint relative invariant* for the Cartesian product action with respect to the determinantal multiplier of weight k. In the case of projective coordinates, where $J\colon \mathcal{P}^{(n)} \times \mathbb{CP}^2 \to \mathbb{C}$, one replaces the determinantal multiplier by a suitable fundamental multiplier of the form (4.22) for appropriate values of k, the weight, and n, the degree, of the covariant.

The simplest example is the "tautologous" function

$$\mathcal{Q}(\mathbf{a}, \mathbf{x}) = \sum_{i=0}^{n} \binom{n}{i} a_i \, x^i y^{n-i}, \tag{4.26}$$

which defines a function $\mathcal{Q}\colon \mathcal{P}^{(n)} \times \mathbb{C}^2 \to \mathbb{C}$. In other words, $\mathcal{Q}(\mathbf{a}, \mathbf{x}) = Q(\mathbf{x})$ coincides with the original binary form, now considered as a function of both its coefficients and its arguments. The function $\mathcal{Q}$ forms a relative invariant whose weight equals the original weight of the binary form: $\mathcal{Q}(\overline{\mathbf{a}}, \overline{\mathbf{x}}) = (\alpha\delta - \beta\gamma)^{-m} \, \mathcal{Q}(\mathbf{a}, \mathbf{x})$. In practice, we shall revert to our old notation and, at the slight risk of confusion, identify the function $\mathcal{Q}$ with the (general) binary form Q.

A fixed point for a matrix multiplier representation will form a vector-valued relative invariant. These are much less common in classical invariant theory, but the following particular example plays an important role in establishing the symbolic method presented in the next chapter.

Exercise 4.42. Let G be any transformation group acting on $\mathbb{R}^n$. Prove that if $I(x)$ is any sufficiently smooth absolute invariant, then its gradient $\nabla I\colon \mathbb{R}^n \to \mathbb{R}^n$ defines a vector-valued relative invariant for the Jacobian multiplier representation of Exercise 4.39. (In differential geometric terms, this means that the Jacobian multiplier governs the induced action of a transformation group on the space of differential one-forms, cf. [**169**].)

An important case occurs when $G = \mathrm{GL}(2)$ acts linearly on $\mathbb{R}^2$, and so the Jacobian multiplier coincides with the identity representation. Of course, there are no nonconstant absolute invariants, and so this case is not particularly interesting. However, consider the Cartesian product action on $\mathcal{P}^{(n)} \times \mathbb{R}^2$, and suppose $J(\mathbf{a}, \mathbf{x})$ is any absolute covariant. Then

$\nabla J = (J_x, J_y)^T$ will be a vector-valued relative invariant for the identity multiplier; in other words,

$$\alpha \overline{J}_{\bar{x}} + \gamma \overline{J}_{\bar{y}} = J_x, \qquad \beta \overline{J}_{\bar{x}} + \delta \overline{J}_{\bar{y}} = J_y. \tag{4.27}$$

Therefore, the gradient operator satisfies

$$\nabla \longmapsto A^{-T} \nabla = \overline{\nabla}, \qquad \text{under the map} \qquad \mathbf{x} \longmapsto \overline{\mathbf{x}} = A\mathbf{x}, \tag{4.28}$$

which indicates that it transforms like a covariant vector! Of course, one can establish this formula by a simple direct calculation; nevertheless, it is of interest to understand how it fits into the general theory of relative invariants and multiplier representations. More generally, if J is a covariant of weight k, then ∇J will be a relative invariant for the matrix-valued multiplier $\mu(A) = (\det A)^k \cdot A$ corresponding to the tensor product of the k^{th} power of the determinantal representation with the identity representation.

Exercise 4.43. Determine what the gradient operator looks like in projective coordinates. What is the appropriate matrix multiplier?