# Contents

**Introduction** iv

1 **Transformations of the Plane** 1
   1.1 Linear transformations ........................................ 2
   1.2 Isometries ................................................. 6
   1.3 Orthogonal linear transformations ............................ 7
   1.4 Affine transformations ....................................... 12
   1.5 Classification of isometries of the plane .................... 16
   1.6 Algebraic automorphisms ..................................... 18
   1.7 Exercises .................................................. 19

2 **Parametric Plane Curves** 20
   2.1 Reparametrization ............................................ 21
   2.2 Length and the arc length parametrization .................. 22
   2.3 Curvature and the Frenet formulas ........................... 23
   2.4 Plane curves up to isometry ................................ 25
   2.5 Exercises .................................................. 26

3 **Affine Algebraic Plane Curves** 28
   3.1 Polynomials and zero loci ................................... 29
   3.2 Smooth and singular points ................................... 32
   3.3 Change of variables and affine equivalence ................. 33
   3.4 Five points determine a conic ............................... 36
   3.5 Plane conics up to affine equivalence ....................... 38
   3.6 Invariants of affine conics ................................ 40
   3.7 Exercises .................................................. 43

4 **Projective Spaces** 45
   4.1 Motivation and construction ................................... 45
   4.2 Linear subspaces of projective space ........................ 48
   4.3 Real and complex projective spaces ............................ 48
   4.4 Projective transformations .................................... 50
4.5 Exercises .......................................................... 53

5 Projective Plane Curves ............................................ 55
  5.1 Homogeneous polynomials and zero loci .................. 56
  5.2 Smooth and singular points .................................. 57
  5.3 Linear change of variables and projective equivalence .. 58
  5.4 Lines and linear functionals .................................. 59
  5.5 Invariants of affine conics revisited ....................... 61
  5.6 Classification of projective conics ......................... 62
  5.7 The cross ratio ................................................ 67
  5.8 Intersection multiplicity .................................... 69
  5.9 The theorems of Pascal, Pappus, and Desargues ........ 70
  5.10 Cubic curves and the group law ............................ 72
  5.11 Nodal and cuspidal cubics .................................. 76
  5.12 Exercises ..................................................... 78

6 Appendix .......................................................... 82
  6.1 Groups and group actions .................................... 82
  6.2 Fields and division rings ...................................... 84
  6.3 The Implicit Function Theorem ............................... 86
  6.4 Exercises ..................................................... 87

Bibliography ........................................................ 88
Introduction

These notes are meant to serve as the text for an undergraduate course in elementary projective geometry. The current notes were written to accompany the course of the same title given at Boğaziçi University in the fall of 2014. Some parts of these notes were recycled from notes I wrote to accompany a course called *Fundamental Problems of Geometry*, which I taught at Brown University in the spring of 2012. Both courses were intended for junior and senior mathematics majors. The second incarnation of this course was particularly aimed at students seeking a career in mathematics instruction. To this end, I made considerable effort to try to “survey” a variety of basic geometric material, with a focus on the geometry of the plane.

In principle linear algebra is the only formal prerequisite, though there are many places where I find it convenient to use the language of group theory, though no actual group theory is really used—it is enough to know the definition of a group, a subgroup, and perhaps a normal subgroup. Also, the idea of a group acting on a set arises (at least implicitly) at many points, so it might be helpful to have an idea what this means—it is a completely elementary concept. Both times I taught the course, I considered devoting a lecture to the basic notions of group theory and group actions (orbits and stabilizers, in particular), but in the end I just mentioned the necessary concepts as they arose, or relegated them to the exercises—that seemed sufficient. At some point I may add an appendix to these notes covering this material.

Let me make some remarks about “level of generality” and the like, mostly intended for the instructor. The usual issues about fields are skirited in these notes in much the same way they are skirited in a typical linear algebra class; that is, we pretty much think of a field as either $\mathbb{R}$ or $\mathbb{C}$ and we don’t much emphasize the difference even between these two fields, except that we can’t get away with this quite as long as we can in linear algebra because we will quickly consider non-linear equations like $x^2 + 1 = 0$ which have radically different behaviour over these two fields. For the sake of exposition, I make most statements over $\mathbb{R}$ (instead of writing $F$ or some such thing to denote a general field), but most of these are true over an arbitrary field, except where (I think) it should be reasonably clear from context that the particular field considered is of central importance to the statement of the result. In fact, I stubbornly insisted on using $\mathbb{R}$ as the “base field” throughout most of the notes, even though most of the “algebro-geometric” content of the notes (of which there is very little, by the way) would probably be easier and more
natural over $\mathbb{C}$.

Although projective geometry and, in particular, the projective plane $\mathbb{R}P^2$, are the main subject matter of these notes, a large part of the text is actually devoted to various geometric considerations in the usual “affine” plane $\mathbb{R}^2$. Without some of this “background” material, much of the projective geometry would seem unmotivated. I also wanted to emphasize the interplay of several different points of view on this subject matter coming from linear algebra, differential geometry, algebraic geometry, and classical axiomatic geometry.

Chapter 1 is devoted to defining and studying various transformation groups, such as the group of invertible linear transformations, the group of isometries, and the group of affine transformations. The latter, in particular, is important because of its relationship with the group of projective transformations (the projective general linear group), which we discuss in the chapter on projective geometry. The material from this chapter is used throughout the rest of the text.

In the first incarnation of the course, I didn’t say anything about isometries, but the group of isometries is so closely related to the group of affine transformations that it seemed strange to discuss one but not the other. Having discussed isometries in the Projective Geometry course, I couldn’t resist fleshing out this material to show how the group of isometries is used in a “real life” geometric study: To this end, I have included Chapter 2 on parametric curves in the plane. I thought it might be useful for pedagogical reasons to have the students return to a study of parametric curves, familiar from calculus courses, equipped with an understanding of the meaning of isometry and the group of isometries of the plane, as the idea of isometry is certainly lurking implicitly in the treatments of parametric curves one sees in any elementary calculus text! To make sure that this chapter contains some new and interesting content, I have included a proof of the (fairly simple) fact that two parametric curves in the plane are related by an orientation preserving isometry iff they have the same length and curvature. My treatment of this is taken from Do Carmo’s book *Differential Geometry of Curves and Surfaces*.

The final two chapters consist of some elementary algebraic geometry of affine and projective plane curves. We introduce the general projective space $\mathbb{R}P^n$, but focus almost exclusively on $\mathbb{R}P^2$. We define the zero locus $Z(f)$ of a polynomial $f$ (or, rather, a homogeneous polynomial in the projective setting), and what it means to be a singular points of $Z(f)$. We explain what it means for polynomials to be “affine equivalent” and for homogeneous polynomials to be “projectively equivalent.” To make this more concrete, we mention the classification of degree two polynomials in two variables up to affine equivalence, though we only give a complete proof of the projective analog of this classification. At the very end we give a brief study of cubic curves in $\mathbb{R}P^2$, giving at least the rough idea of the group law on the set of smooth points of an irreducible cubic.
Chapter 1

Transformations of the Plane

Let us agree that a transformation (of the plane) is a bijective function $f : \mathbb{R}^2 \to \mathbb{R}^2$. Among all transformations we can single out various classes of transformations preserving various additional structures possessed by $\mathbb{R}^2$. In this chapter we will define and study the following types of transformations:

1. homeomorphisms
2. invertible linear transformations
3. isometries
4. orthogonal linear transformations
5. affine transformations
6. algebraic automorphisms

One can consider these types of transformations more generally in the case of $\mathbb{R}^n$, or over other fields. Instead of $\mathbb{R}^n$, one could work with an abstract vector space, or with a vector space equipped with some extra structure. We will make some remarks about these more general settings, but our main focus will be on the case $\mathbb{R}^2$, as this will be most important in the later chapters. The set of transformations of each type forms a group. In some cases, we will say something about the structure of this group.

Although detailed definitions will be given in the sections of this chapter, we will give a summary now for the convenience of the reader. A homeomorphism is a continuous bijection $f : \mathbb{R}^2 \to \mathbb{R}^2$ with continuous inverse $f^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$. Homeomorphisms are the most general type of transformation that will be of any use to us—every other type of transformation we consider will be a homeomorphism. Homeomorphisms preserve properties of subsets of $\mathbb{R}^2$ such as connectedness, being open, or being compact (closed and bounded).

An invertible linear transformation is a bijection $f : \mathbb{R}^2 \to \mathbb{R}^2$ that “commutes with scalar multiplication and vector addition.” The importance of invertible linear transformations results mainly from their usage in linear algebra. An isometry of $\mathbb{R}^2$ is a bijection $f : \mathbb{R}^2 \to \mathbb{R}^2$ that preserves distance. Isometries can be useful when studying
various geometric properties of subsets of $\mathbb{R}^2$ such as lengths, areas, angles, and so forth. One can define an **orthogonal linear transformation** to be a map $f : \mathbb{R}^2 \to \mathbb{R}^2$ which is both an isometry and a linear transformation. We will describe all of these. An **affine transformation** is a map $f : \mathbb{R}^2 \to \mathbb{R}^2$ expressible as a composition of a linear transformation and a translation. Affine transformations arise naturally when studying more “algebraic” subsets of $\mathbb{R}^2$ (zero loci of polynomials) and their properties. They are also closely related to the projective transformations ($\S$4.4) that we will study when we introduce projective geometry in Chapter 4. An **algebraic automorphism** is a bijection $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that the coordinates of both $f$ and its inverse are given by polynomial functions of two variables. These are the most subtle kind of transformation that we will consider—we won’t have too much to say about them, but it would be amiss not to mention these along with the others!

One reason to be interested in these sorts of transformations is the following: If one is interested in calculating some quantity $Q(A)$ associated to a subset $A \subseteq \mathbb{R}^2$ (or perhaps to several subsets $A_1, \ldots, A_n \subseteq \mathbb{R}^2$), it is often convenient to calculate the corresponding quantity $Q(f(A))$ for $f(A)$ for some transformation $f$ for which it is known *a priori* that $Q(A) = Q(f(A))$. Of course this seems silly when expressed in such abstract terms, but imagine, for example, the problem of computing the circumference of a circle in $\mathbb{R}^2$, or, more generally, the length $\ell(A)$ of some other curve $A$ in $\mathbb{R}^2$. Because this problem involves doing significant computation (integration), it is usually helpful to move $A$ to some other position via an isometry $f$ (for example: move the circle so it is centered at the origin), which won’t effect the result of the computation (because one knows “a priori” that length is isometry invariant), but which will make the computation easier to actually carry out. Similar considerations are common in linear algebra, where various operations are naturally defined and studied in one basis; one then uses invertible linear transformations to re-express things in the standard basis. We will make use of the various kinds of transformations for purposes such as this in the other chapters of the book.

### 1.1 Linear transformations

Although I assume the student has some background in linear algebra, let me briefly review some of the more relevant material. Throughout, we work with a fixed **field** $\mathbb{K}$, often called the base field. The reader unfamiliar with the general notion of a field should simply focus on the case where $\mathbb{K}$ is the field $\mathbb{Q}$ of rational numbers, the field $\mathbb{R}$ of real numbers, or the field $\mathbb{C}$ of complex numbers. Let

$$\mathbb{K}^n := \{(x_1, \ldots, x_n) : x_1, \ldots, x_n \in \mathbb{K}\}$$

be the set of ordered $n$-tuples of elements of $\mathbb{K}$. We will use symbols like $x$, $y$, etc. to denote elements of $\mathbb{K}^n$, writing $x_1, \ldots, x_n$ for the coordinates of $x$.

From the point of view of linear algebra, the set $\mathbb{K}^n$ is equipped with two important structures: **(vector) addition** and **scalar multiplication**. For $x, y \in \mathbb{K}^n$, we let

$$x + y := (x_1 + y_1, \ldots, x_n + y_n)$$
be the “vector sum” of $x$ and $y$. Given $\lambda \in \mathbb{K}$ and $x \in \mathbb{K}^n$, we set
\[
\lambda x := (\lambda x_1, \ldots, \lambda x_n).
\]
The vector $\lambda x$ is called a **scalar multiple** of $x$; the term **rescaling of** $x$ is reserved for a vector of the form $\lambda x$ with $\lambda \in \mathbb{K}^* := \mathbb{K} \setminus \{0\}$. Vector addition and scalar multiplication satisfy various properties, which one is led to investigate and “axiomatize”: for example, scalar multiplication “distributes over vector addition.” After thinking about these properties, one is eventually led to define a **vector space** (or a **vector space over** $\mathbb{K}$ if $\mathbb{K}$ is not clear from context) to be a set $V$ equipped with two functions, written
\[
\begin{align*}
V \times V &\to V \\
(u, v) &\mapsto u + v \\
\mathbb{R} \times V &\to V \\
(\lambda, v) &\mapsto \lambda v,
\end{align*}
\]
and called, respectively, **(vector) addition** and **scalar multiplication** satisfying various axioms that we will not repeat here (see any linear algebra textbook).

It is customary in linear algebra to write the elements $x$ of $\mathbb{K}^n$ as column vectors, but in most of the inline text of these notes I will usually write these as row vectors $x = (x_1, \ldots, x_n)$ for aesthetic reasons. I could put some kind of “transpose” notation in to be precise, but I think it will always be clear when I am using row vectors or column vectors. I usually use $u$, $v$, etc. for elements of an abstract vector space, which I usually denote $U$, $V$, or $W$. In linear algebraic context, I try to use Greek letters like $\lambda$, $\mu$, and so forth for elements of $\mathbb{K}$, also called **scalars**.

**Definition 1.1.1.** If $V$ and $W$ are vector spaces, a **linear transformation** from $V$ to $W$ is a function $f : V \to W$ satisfying $f(\lambda v) = \lambda f(v)$ and $f(u + v) = f(u) + f(v)$ for every $\lambda \in \mathbb{K}$, $u, v \in V$. Informally, one says that $f$ “respects addition and scalar multiplication.” Let $\text{Hom}_\mathbb{K}(V, W)$ be the set of linear transformations $f : V \to W$.

It is easy to check that a composition of linear transformations is a linear transformation, so that usual composition of functions defines a “composition of linear transformations”
\[
\text{Hom}_\mathbb{K}(V, W) \times \text{Hom}_\mathbb{K}(U, V) \to \text{Hom}_\mathbb{K}(U, W)
\]
\[
(f, g) \mapsto fg
\]
for vector spaces $U, V, W$. It is also easy to see that, if $f : V \to W$ is a bijective linear transformation, then its inverse $f^{-1} : W \to V$ is also a (bijective) linear transformation.

**Definition 1.1.2.** An $m \times n$ **matrix** $A = (A_{i,j})$ consists of elements $A_{i,j} \in \mathbb{K}$, called the **entries** of $A$, one for each
\[
(i, j) \in \{1, \ldots, m\} \times \{1, \ldots, n\}.
\]
We write $\text{Mat}(m \times n)$ for the set of $m \times n$ matrices; if there is any ambiguity about $\mathbb{K}$, then we write $\text{Mat}(m \times n, \mathbb{K})$. 

One thinks of the entries $A_{i,j}$ of a matrix $A$ as being arranged in a rectangular array such that $A_{i,j}$ is in row $i$ and column $j$. Thus, for example, a $2 \times 3$ matrix $A$ is written in the form

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ A_{2,1} & A_{2,2} & A_{2,3} \end{pmatrix}.$$  

It is common to drop the comma in the subscript, writing $A_{ij}$ instead of $A_{i,j}$, though this is a bit sloppy since “$ij$” might be mistaken for the product of $i$ and $j$. The entries of a $2 \times 2$ matrix are almost always called $a, b, c, d$ rather than $A_{1,1}, A_{1,2}, A_{2,1},$ and $A_{2,2}$, so that a typical $2 \times 2$ matrix is written

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \quad (1.1)$$

The fundamental operation with matrices is the (matrix) product: Given an $m \times n$ matrix $A$ and an $n \times k$ matrix $B$, one defines an $m \times k$ “product” matrix $AB$ by setting

$$(AB)_{i,j} := \sum_{p=1}^{n} A_{i,p} B_{p,j}.$$ 

To each square matrix $A \in \text{Mat}(n \times n)$, one can associate a real number $\det A$, called the determinant of $A$. If $A$ is $1 \times 1$, one not surprisingly has $\det A = A_{1,1}$ equal to the unique entry of $A$. If $A$ is a $2 \times 2$ matrix as in $(1.1)$, then $\det A = ad - bc$. In general, in an elementary linear algebra class, one defines $\det A$ by some kind of inductive procedure (“expansion by minors”), though there are more satisfying ways of doing this. In any case, the most important property of the determinant is that

$$\det(AB) = (\det A)(\det B)$$

for all $n \times n$ matrices $A, B$.

The “matrix product” is motivated by linear algebraic considerations as follows:

For $i \in \{1, \ldots, n\}$, let $e_i \in \mathbb{K}^n$ be the point with $i^{th}$ coordinate $1$ and all other coordinates zero. For example, $e_1 = (1, 0)$ and $e_2 = (0, 1)$ in $\mathbb{K}^2$. If $f : \mathbb{K}^n \to \mathbb{K}^m$ is a linear transformation, we define an $m \times n$ matrix $A = A_f$ by letting the $i^{th}$ column ($i = 1, \ldots, n$) of $A$ be $f(e_i) \in \mathbb{K}^n$ (viewing this as a column vector). For example, if $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $f(e_1) = (a, c), f(e_2) = (b, d)$, then

$$A_f = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$ 

Given an $m \times n$ matrix $A$, we define a function $f_A : \mathbb{K}^n \to \mathbb{K}^m$ by letting $f_A(x)$ be the matrix product $Ax$ defined above, viewing $x$ as a column vector ($n \times 1$ matrix). The following standard result of linear algebra says that “matrices are the same thing as linear transformations and matrix multiplication is the same thing as composition:”
Proposition 1.1.3. The maps \( f \mapsto A_f \) and \( A \mapsto f_A \) defined above are inverse bijections between \( \text{Hom}_\mathbb{K}(\mathbb{K}^n, \mathbb{K}^m) \) and \( \text{Mat}(m \times n) \). Under these bijections, matrix multiplication corresponds to composition of linear transformations

\[
\text{Mat}(m \times n) \times \text{Mat}(n \times k) \to \text{Mat}(m \times k)
\]

That is, \( f_{AB} = f_A f_B \) and \( A_f g = A_f A_g \).

Recall the following standard result of linear algebra, which, among other things, explains the importance of the determinant:

Proposition 1.1.4. For a linear transformation \( f : \mathbb{K}^n \to \mathbb{K}^n \), the following are equivalent:

1. \( f(e_1), \ldots, f(e_n) \) is a basis for \( \mathbb{K}^n \).
2. \( f \) is surjective.
3. \( f \) is injective.
4. \( f \) is bijective.
5. \( \text{Ker } f = \{0\} \).
6. There is a linear transformation \( g : \mathbb{K}^n \to \mathbb{K}^n \) such that \( gf(x) = fg(x) = x \) for all \( x \in \mathbb{R}^n \).
7. The determinant \( \det A_f \) of the matrix \( A_f \) associated to \( f \) is non-zero.

Definition 1.1.5. A linear transformation \( f : \mathbb{K}^n \to \mathbb{K}^n \) satisfying the equivalent conditions of the above proposition is called an invertible linear transformation.

It is clear that a composition of invertible linear transformations is an invertible linear transformation. The set of invertible linear transformations \( f : \mathbb{K}^n \to \mathbb{K}^n \) forms a group under composition, denoted \( \text{GL}_n(\mathbb{K}) \), and often called the general linear group. Blurring the distinction between a linear transformation and the corresponding matrix, an element of \( \text{GL}_2(\mathbb{K}) \) can be viewed as a \( 2 \times 2 \) matrix

\[
A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}
\]

where \( \det A = ad - bc \neq 0 \).
### 1.2 Isometries

Recall that a **metric** $d$ on a set $X$ is a function

$$d : X \times X \to \mathbb{R}$$

satisfying the properties:

1. $d(x, y) \geq 0$ for all $x, y \in X$ with equality iff $x = y$.
2. $d(x, y) = d(y, x)$ for all $x, y \in X$.
3. $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

A **metric space** $(X, d)$ is a set $X$ equipped with a metric $d$. We usually just write $X$ to denote a metric space, leaving $d$ implicit; we often write $d_X$ for the metric on $X$ if there is any chance of confusion (for example, if we are considering more than one metric space).

For example, it is not hard to check that

$$d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$$

$$d(x, y) := \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

is a metric on $\mathbb{R}^n$, called the **standard metric**. By default, we regard $\mathbb{R}^n$ as a metric space by using the standard metric unless some other metric is explicitly mentioned. In general, if $d$ is a metric on a set $X$ and $A$ is a subset of $X$, then by restriction $d$ determines a metric on $A$, also abusively denoted $d$. In particular, the standard metric on $\mathbb{R}^n$ determines a metric on any subset $A \subseteq \mathbb{R}^n$; we always use this metric to view $A$ as a metric space, unless explicitly mentioned to the contrary.

One defines continuity for maps between metric spaces by using the familiar “epsilon-delta definition:”

**Definition 1.2.1.** A function $f : X \to Y$ between metric spaces $X, Y$ is called **continuous** iff for every $x \in X$ and every positive real number $\epsilon$, there is a positive real number $\delta = \delta(x, \epsilon)$ such that $d_Y(f(x), f(x')) < \epsilon$ whenever $x'$ is a point of $X$ with $d_X(x, x') < \delta$.

A **homeomorphism** $f : X \to Y$ is a continuous bijection whose inverse $f^{-1} : Y \to X$ is also continuous.

It is easy to see that a composition of continuous maps between metric spaces is continuous, hence a composition of homeomorphisms is a homeomorphism. In particular, the set $\text{Homeo}(\mathbb{R}^n)$ of homeomorphisms $f : \mathbb{R}^n \to \mathbb{R}^n$ is a group under composition. It is also not hard to check that any linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuous. It follows that the group $\text{GL}_n(\mathbb{R})$ of invertible linear transformations is a subgroup of $\text{Homeo}(\mathbb{R}^n)$.

**Definition 1.2.2.** A function $f : X \to Y$ between metric spaces $X$ and $Y$ is called **distance preserving** iff $d_X(x, x') = d_Y(f(x), f(x'))$ for all $x, x' \in X$. An **isometry** $f : X \to Y$ is a distance preserving bijection.
The following facts are readily established:

1. A composition of distance preserving functions is distance preserving.

2. A distance preserving function is continuous (one can take $\delta = \epsilon$) and one-to-one (injective).

3. The inverse of an isometry is an isometry.

4. Every isometry is a homeomorphism.

5. A composition of isometries is an isometry.

In particular, the set $\text{Isom}(\mathbb{R}^n)$ of isometries $f : \mathbb{R}^n \to \mathbb{R}^n$ is a subgroup of the group $\text{Homeo}(\mathbb{R}^n)$.

The definitions above yield notions of continuity, isometry, etc. for functions between subsets of $\mathbb{R}^n$. We will see later that any distance preserving function $\mathbb{R}^2 \to \mathbb{R}^2$ is automatically bijective (i.e. is an isometry). (This is true, more generally, for $\mathbb{R}^n$.) Indeed, we will completely classify the isometries of $\mathbb{R}^2$, and this will fall out as a consequence—the reader may wish to try to directly prove this assertion.

**Example 1.2.3. (Translations)** Let $\mathbb{K}$ be a field. Fix $t \in \mathbb{K}^n$. Define a function $f_t : \mathbb{K}^n \to \mathbb{K}^n$ by $f_t(x) := t + x$. When $\mathbb{K} = \mathbb{R}$, the function $f_t : \mathbb{R}^n \to \mathbb{R}^n$ preserves distance because

$$
d(f_t(x), f_t(y)) = \sqrt{\sum_{i=1}^{n} ((t_i + x_i) - (t_i + y_i))^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} = d(x, y).
$$

A function $f : \mathbb{K}^n \to \mathbb{K}^n$ equal to $f_t$ for some $t \in \mathbb{K}^n$ is called a translation—the $t$ will be unique since we can recover $t$ from $f_t$ by the obvious formula $t = f_t(0)$. Notice that the composition $f_s f_t$ of two translations is the translation $f_{s+t}$. The set of translations $\mathbb{K}^n \to \mathbb{K}^n$ is a group under composition of functions; this group is isomorphic to $\mathbb{K}^n$ (under vector addition) via the maps $t \mapsto f_t$, $f \mapsto f(0)$. The inverse of the translation $f_t$ is the translation $f_{-t}$. In particular, every translation $f_t : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry.

### 1.3 Orthogonal linear transformations

Given a real vector space $V$ (vector space over the field of real numbers $\mathbb{R}$), recall from linear algebra that an inner product (more precisely: positive definite inner product) on $V$ is a function $V \times V \to \mathbb{R}$, usually written $(u, v) \mapsto \langle u, v \rangle$, satisfying the conditions

1. $\langle u, v \rangle = \langle v, u \rangle$
2. \( \langle u, u \rangle \geq 0 \) with equality iff \( u = 0 \)

3. \( \langle u + u', v \rangle = \langle u, v \rangle + \langle u', v \rangle \)

4. \( \langle \lambda u, v \rangle = \lambda \langle u, v \rangle \)

for all \( u, u', v \in V, \lambda \in \mathbb{R} \).

For example,

\[
\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}
\]

\[
\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i
\]

defines an inner product on \( \mathbb{R}^n \), called the \textit{standard inner product}.

An \textit{inner product space} is a real vector space \( V \) equipped with an inner product. An inner product \( \langle \cdot, \cdot \rangle \) on \( V \) gives rise to a metric \( d \) on \( V \) defined by

\[
d(u, v) := \langle u - v, u - v \rangle
\]

Two vectors \( u, v \) in an inner product space \( V \) are called \textit{orthogonal} iff \( \langle u, v \rangle = 0 \). More generally, one defines the angle \( \theta \in [0, \pi] \) between two non-zero vectors \( u, v \) in \( V \) by the formula

\[
\cos \theta = \frac{\langle u, v \rangle}{|u||v|}
\]

where the symbol

\[
|u| := d(u, 0)
\]

\[
= \sqrt{\langle u, u \rangle}
\]

denotes the \textit{magnitude} of \( u \). (One checks that the right hand side of (1.2) is in \([-1, 1]\) so that this makes sense.)

Unless mentioned to the contrary, we always regard \( \mathbb{R}^n \) as an inner product space by equipping it with the standard inner product. The standard inner product on \( \mathbb{R}^n \) gives rise to the standard metric on \( \mathbb{R}^n \). Let us check that the notion of the angle between two vectors defined by the standard inner product is the usual angle in \( \mathbb{R}^n \). Recall that the Law Of Cosines says that in any triangle with sides of length \( a, b, c \), we have

\[
c^2 = a^2 + b^2 - 2ab \cos C,
\]

where \( C \) is the angle opposite the side of length \( c \). If we apply this to the “triangle” whose vertices are \( 0 \), and two non-zero vectors \( x, y \in \mathbb{R}^n \), letting \( c \) be the side connecting \( x \) to \( y \), we find

\[
\langle u, v \rangle = |u|^2 + |v|^2 - 2|u||v| \cos \theta.
\]

If we expand out the inner product on the left, cancel the \( |u|^2 + |v|^2 \) appearing on both sides and divide by \(-2\), we find that the angle \( \theta \) between \( u \) and \( v \) is the same as the
“abstractly defined” angle $\theta$. In particular, $x, y \in \mathbb{R}^n$ are orthogonal in the usual sense iff

$$\langle x, y \rangle = 0.$$

One says that a list of $n$ vectors $v_1, \ldots, v_n$ in an $n$-dimensional inner product space $V$ is an orthonormal basis for $V$ iff $\langle v_i, v_j \rangle = \delta_{i,j}$ is zero when $i \neq j$ and one when $i = j$ for all $i, j \in \{1, \ldots, n\}$. An orthonormal basis is, in particular, a basis: By linear algebra, it suffices to check that the vectors $v_1, \ldots, v_n$ are linearly independent. Indeed, if we have $\sum_{i=1}^n \lambda_i v_i = 0$, then taking $\langle \cdot, v_j \rangle$ and using properties of the inner product and the orthogonality assumption on the $v_i$, we see that $\lambda_j = 0$.

The standard inner product on $\mathbb{R}^n$ gives rise to a nice relationship between the linear algebra of $\mathbb{R}^n$ and the isometries of $\mathbb{R}^n$:

**Proposition 1.3.1.** For a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^n$, the following are equivalent:

1. $\langle f(x), f(y) \rangle = \langle x, y \rangle$ for all $x, y \in \mathbb{R}^n$.

2. The columns of the matrix $A = A_f$ associated to $f$ form an orthonormal basis for $\mathbb{R}^n$.

3. $f$ preserves distance.

4. $f$ is an isometry of $\mathbb{R}^n$.

**Proof.** To see that (1) implies (2), just note that the $i^{th}$ column of $A_f$ is $f(e_i)$, so we have to show that $\langle f(e_i), f(e_j) \rangle = \delta_{i,j}$. But by (1) we have

$$\langle f(e_i), f(e_j) \rangle = \langle e_i, e_j \rangle$$

and it is clear from the definition of the standard inner product on $\mathbb{R}^n$ that this is $\delta_{i,j}$.

To see that (2) implies (1), write $x = \sum_{i=1}^n x_i e_i$, $y = \sum_{i=1}^n y_i e_i$ for all real numbers $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. Then $f(x) = \sum_{i=1}^n x_i f(e_i) = \sum_{i=1}^n A_i$, where $A_i = f(e_i)$ is the $i^{th}$ column of $A = A_f$. Using this, the analogous formula for $f(y)$, the hypothesis on $A$, and properties of $\langle \cdot, \cdot \rangle$, we compute

$$\langle f(x), f(y) \rangle = \left\langle \sum_{i=1}^n x_i A_i, \sum_{i=1}^n y_i A_i \right\rangle$$

$$= \sum_{i,j=1}^n x_i y_j \langle A_i, A_j \rangle$$

$$= \sum_{i,j=1}^n x_i y_j \delta_{i,j}$$

$$= \sum_{i=1}^n x_i y_j$$

$$= \langle x, y \rangle.$$
To see that (1) implies (3), we compute, for any \( x, y \in \mathbb{R}^n \) that
\[
d(f(x), f(y)) = \langle f(x) - f(y), f(x) - f(y) \rangle \\
= \langle f(x - y), f(x - y) \rangle \\
= \langle x - y, x - y \rangle \\
= d(x, y)
\]
using linearity of \( f \) for the second equality and (1) for the third equality.

To see that (3) implies (2), let \( A_i = f(e_i) \) denote the \( i \)th column of \( A = Af \). First
note that
\[
\langle A_i, A_i \rangle = \langle A_i - 0, A_i - 0 \rangle \\
= d(A_i, 0) \\
= d(f(e_i), f(0)) \\
= d(e_i, 0),
\]
so we see that each column of \( A \) has magnitude one. We still have to show that any two
distinct columns \( A_i, A_j \) of \( A \) are orthogonal, so fix distinct \( i, j \in \{1, \ldots, n\} \) and let \( \theta \) be
the angle between \( A_i \) and \( A_j \). Since \( f \) preserves distance, we have
\[
d(A_i, A_j) = d(f(e_i), f(e_j)) \\
= d(e_i, e_j) \\
= \sqrt{2}.
\]

Now the Law Of Cosines gives
\[
2 = d(A_i, A_j)^2 \\
= |A_i|^2 + |A_j|^2 - 2|A_i||A_j| \cos \theta \\
= 2 - 2 \cos \theta,
\]
so we must have \( \cos \theta = 0 \), hence \( A_i \) and \( A_j \) are orthogonal, as desired.

To see that (3) implies (4), we just need to show that a linear transformation \( f : \mathbb{R}^n \to \mathbb{R}^n \) which preserves distance is surjective. (Recall that every distance preserving map is
injective and that an isometry is a distance preserving bijection.) Since \( f \) is injective,
we can just use the general linear algebra fact that an injective linear transformation
between vector spaces of the same (finite) dimension is bijective. Alternatively, we could
use the fact that (3) implies (2) to establish bijectivity of \( f \).

Obviously (4) implies (3). \( \square \)

**Definition 1.3.2.** A linear transformation \( f : \mathbb{R}^n \to \mathbb{R}^n \) satisfying the equivalent con-
ditions of Proposition 1.3.1 is called an **orthogonal linear transformation**, or simply an
**orthogonal transformation**. Similarly, a matrix \( A \in \text{Mat}(n \times n) \) is called an **orthogonal ma-
trix** iff the corresponding linear transformation \( f_A \) is an orthogonal linear transformation
(iff the columns of \( A \) form an orthonormal basis for \( \mathbb{R}^n \)).
Since a composition of linear transformations is a linear transformation and a composition of isometries is an isometry, we see that a composition of orthogonal linear transformations is an orthogonal linear transformation. The group of orthogonal linear transformations $\mathbb{R}^n \to \mathbb{R}^n$ is denoted $O_n$. The group $O_n$ may be viewed as a subgroup of both $\text{GL}_n(\mathbb{R})$ and $\text{Isom}(\mathbb{R}^n)$. Indeed, the proposition above shows that

$$O_n = \text{GL}_n(\mathbb{R}) \cap \text{Isom}(\mathbb{R}^n),$$

viewing all of these groups as subgroups of $\text{Homeo}(\mathbb{R}^n)$.

The orthogonal transformations of $\mathbb{R}^n$ can be explicitly described:

**Proposition 1.3.3.** Suppose $f : \mathbb{R}^2 \to \mathbb{R}^2$ is an orthogonal transformation, with associated matrix $A = A_f$. Then the determinant of $A$ is $\pm 1$. If $\det A = 1$, then there is a unique angle $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ such that

$$A = \begin{pmatrix} \cos \theta - \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

If $\det A = -1$, then there is a unique angle $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ such that

$$A = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

**Proof.** If $f$ is an orthogonal transformation, then $f(e_1)$ and $f(e_2)$ (the columns of $A$) have magnitude one, hence lie on the unit circle

$$S^1 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\},$$

so we can write $f(e_1) = (\cos \theta, \sin \theta)$ for a unique $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Since $f(e_2)$ must also be a point of the unit circle orthogonal to $f(e_1)$, the second column of $A$ is either

$$(\cos(\theta + \pi/2), \sin(\theta + \pi/2)) = (-\sin \theta, \cos \theta)$$

or

$$(\cos(\theta - \pi/2), \sin(\theta - \pi/2)) = (\sin \theta, -\cos \theta).$$

\[\square\]

**Remark 1.3.4.** Proposition 1.3.3 characterizing the orthogonal transformations of $\mathbb{R}^2$ in particular shows that each such orthogonal transformation has determinant $\pm 1$. It is true more generally that any orthogonal transformation of $\mathbb{R}^n$ has determinant $\pm 1$. To see this, observe that for any square matrix $A$ with real entries, we have

$$(A^tA)_{ij} = \langle A_i, A_j \rangle,$$

where $A^t$ denotes the transpose of $A$ and $A_i$ denotes the $i^{th}$ column of $A$. In particular, we see that $A$ is an orthogonal matrix iff $A^tA = \text{Id}$. Since $\det A = \det A^t$ by general theory of determinants, this shows that $(\det A)^2 = \det \text{Id} = 1$. 

Example 1.3.5. (Rotations) If \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is an orthogonal transformation with \( \det A_f = 1 \) (also called a *special orthogonal transformation*) and we write \( f \) as in the above proposition, then, geometrically, \( f \) is given by counter-clockwise rotation around the origin through the angle \( \theta \). Indeed, one checks readily that \( d(0, f(x)) = d(0, x) \) for \( x \in \mathbb{R}^2 \), and that if \( \eta \) is the counter-clockwise angle between the positive \( x \)-axis and \( x \in \mathbb{R}^2 \setminus \{0\} \), then the counter-clockwise angle between the positive \( x \)-axis and \( f(x) \in \mathbb{R}^2 \setminus \{0\} \) is \( \eta + \theta \). The group \( SO_2 \) of special orthogonal transformations is a normal subgroup of \( O_2 \). One checks easily that the map taking \( \theta \in \mathbb{R}/2\pi\mathbb{Z} \) to the first matrix \( A = A_\theta \) in the above proposition (or rather, to the corresponding orthogonal transformation \( f_A \)) defines an isomorphism of groups

\[
(\mathbb{R}/2\pi\mathbb{Z}, +) \cong SO_2.
\]

One can also define an isometry \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) by rotating around an arbitrary point of \( \mathbb{R}^2 \). The reader can check that such an isometry can be expressed as a composition of translations (Example 1.2.3) and a rotation around the origin.

Example 1.3.6. (Reflections) The matrix

\[
A := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
\]

is clearly an orthogonal matrix, so the associated linear transformation \( f = f_A : \mathbb{R}^2 \to \mathbb{R}^2 \) is an orthogonal linear transformation, hence, in particular, an isometry. Note that \( f(x, y) = (x, -y) \), so that \( f \) can be described geometrically by reflecting over the \( x \)-axis. Since

\[
\begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix} = \begin{pmatrix} \cos \theta - \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},
\]

the classification of orthogonal transformations of \( \mathbb{R}^2 \) in Proposition 1.3.3 shows that every such transformation is either a rotation around the origin, or the reflection across the \( x \)-axis followed by such a rotation.

More generally, one can define an isometry \( r_L : \mathbb{R}^2 \to \mathbb{R}^2 \) by reflecting across an arbitrary line \( L \subset \mathbb{R}^2 \). The reader can check that such an isometry can be expressed as a composition of rotations, translations, and the reflection over the \( x \)-axis discussed previously. In particular, it is a standard linear algebra exercise to show that reflection over a line passing through the origin is an orthogonal linear transformation of determinant \(-1\).

1.4 Affine transformations

Let \( \mathbb{K} \) be a field, \( \mathbb{K}^n \) the set of ordered \( n \)-tuples of elements of \( \mathbb{K} \), as in §1.1.

Definition 1.4.1. An *affine transformation* is a function \( f : \mathbb{K}^n \to \mathbb{K}^n \) expressible as the composition of a linear transformation (Definition 1.1.1) followed by a translation (Example 1.2.3).
The first thing to observe is that the expression of an affine transformation \( f = f_t A \) as a linear transformation \( A \) followed by a translation \( f_t \) is unique because we can recover \( t \) from \( f \) via
\[
f(0) = f_t(A(0)) = f_t(0) = t
\]
and then we can recover \( A \) from \( f \) and \( t \) by the obvious formula \( A = f_{-t} f \). In other words, if we fix \( A \in \text{GL}_n(\mathbb{K}) \) and \( t \in \mathbb{K}^n \), then we get an affine transformation \([A, t]: \mathbb{K}^n \to \mathbb{K}^n\) by setting \([A, t](x) := Ax + t\). Every affine transformation is of the form \([A, t]\) for a unique \( A \in \text{GL}_n(\mathbb{K}) \) and \( t \in \mathbb{K}^n \), so that the set of affine transformations is in bijective correspondence with the set \( \text{GL}_n(\mathbb{K}) \times \mathbb{K}^n \).

Affine transformations form a group under composition, denoted \( \text{Aff}(\mathbb{K}^n) \). It is important to understand that, although \( \text{Aff}(\mathbb{K}^n) \) is in bijective correspondence with \( \text{GL}_n(\mathbb{K}) \times \mathbb{K}^n \) via the bijection described above, this bijection is not an isomorphism of groups between \( \text{Aff}(\mathbb{K}^n) \) and the product group \( \text{GL}_n(\mathbb{K}) \times \mathbb{K}^n \). To see why this is the case, let us calculate the composition of two affine transformations \([A, t] \) and \([A', t']\).

Given \( x \in \mathbb{K}^n \), we compute
\[
([A, t] \circ [A', t'])(x) = [A, t](A'x + t') = A(A'x + t') + t = (AA')x + (At' + t),
\]
so the composition of affine transformations is given by
\[
[A, t][A', t'] = [AA', At' + t]. \tag{1.3}
\]

It follows from (1.3) that the inverse of the affine transformation \([A, t] \) is given by
\[
[A, t]^{-1} = [A^{-1}, -A^{-1}t]. \tag{1.4}
\]

If we view \( \text{Aff}(\mathbb{K}^n) \) as the set of pairs \([A, t] \in \text{GL}_n(\mathbb{K}) \times \mathbb{K}^n \) with composition law (1.3), then the group of translations \( \mathbb{K}^n \) is identified with the subgroup of \( \text{Aff}(\mathbb{K}^n) \) consisting of the pairs \([A, t]\) where \( A = \text{Id} \), and the group \( \text{GL}_n(\mathbb{K}) \) of invertible linear transformations is identified with the subgroup of \( \text{Aff}(\mathbb{K}^n) \) consisting of the pairs \([A, t]\) where \( t = 0 \).

**Theorem 1.4.2.** The group of translations \( (\mathbb{K}^n, +) \) is a normal subgroup of \( \text{Aff}(\mathbb{K}^n) \) and every element of \( \text{Aff}(\mathbb{K}^n) \) can be written uniquely as a composition of a translation and an invertible linear transformation. In other words, \( \text{Aff}(\mathbb{K}^n) \) is a semi-direct product of \( (\mathbb{K}^n, +) \) and \( \text{GL}_n(\mathbb{K}) \).

**Proof.** The only statement not proved in the discussion above is the normality of \( (\mathbb{K}^n, +) \). We calculate the conjugate of a translation \([\text{Id}, s]\) by an arbitrary affine transformation \([A, t]\) by using the formulas (1.3) and (1.4):
\[
[A, t][\text{Id}, s][A, t]^{-1} = [A, t][\text{Id}, s][A^{-1}, -A^{-1}t] = [A, As + t][A^{-1}, -A^{-1}t] = [\text{Id}, A(-A^{-1}t) + As + t] = [\text{Id}, As].
\]
This is a translation, as desired. \( \square \)
At present, the consideration of the groups \( \text{Aff}(\mathbb{K}^n) \) may seem somewhat unmotivated. In fact, these groups are perhaps the most important groups that arise in geometry—here are some reasons why:

1. They are used to define and study (affine) linear changes of coordinates and (affine) linear changes of variables, hence they play a role in problems of classifying polynomials and (affine) plane curves.

2. The group \( \text{Aff}(\mathbb{K}^n) \) is closely related to the projective general linear group, and hence to the geometry of the projective space \( \mathbb{K}P^n \) and the inclusion \( \mathbb{K}^n \hookrightarrow \mathbb{K}P^n \), which we shall study later.

3. As we shall see in §1.5, the group of affine transformations \( \text{Aff}(\mathbb{R}^2) \) is closely related to the group of isometries of \( \mathbb{R}^2 \).

4. We will see in Theorem 1.4.5 that affine transformations arise naturally when considering the geometry of lines in \( \mathbb{R}^2 \).

**Proposition 1.4.3.** Suppose \( P_0, \ldots, P_n \in \mathbb{K}^n \) are \((n+1)\) points of \( \mathbb{K}^n \) in general position, meaning they are not all contained in a translate of some linear subspace \( V \subseteq \mathbb{K}^n \) of dimension \(< n \). Then there is a unique affine transformation \([A,t] \in \text{Aff}(\mathbb{K}^n)\) such that \([A,t](0) = P_0\) and \([A,t](e_i) = P_i\) for \( i = 1, \ldots, n \).

**Remark 1.4.4.** When \( n = 2 \), the condition in Proposition 1.4.3 that \( P, Q, R \in \mathbb{R}^2 \) be in “general position” just says that \( P, Q, R \) are not collinear (are not contained in a line).

**Proof.** Since the \( P_i \) are in general position, the vectors \( P_1 - P_0, \ldots, P_n - P_0 \) must be linearly independent, hence by linear algebra, they form a basis for \( \mathbb{K}^n \) and there is a unique linear transformation \( A : \mathbb{K}^n \to \mathbb{K}^n \) with \( Ae_i = P_i - P_0 \). The affine transformation \([A,P_0] \) given by the linear transformation \( A \) followed by translation along \( P_0 \) is clearly the unique affine transformation with the desired properties.

**Theorem 1.4.5.** A function \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is an affine transformation iff \( f \) is a continuous function with the following property (*) : For any \( P, Q, R \in \mathbb{R}^2 \), \( P, Q, R \) are collinear iff \( f(P), f(Q), f(R) \) are collinear.

**Proof.** It is clear that any affine transformation of \( \mathbb{R}^2 \) is a continuous function satisfying (*) and that a composition of continuous functions satisfying (*) is also a continuous function satisfying (*). Suppose now that \( h \) is a continuous function satisfying (*). Then \( h(0), h(e_1) \) and \( h(e_2) \) are non-collinear, so by the previous proposition there is an affine transformation \( g : \mathbb{R}^2 \to \mathbb{R}^2 \) with \( g(0) = h(0), g(e_1) = h(e_1) \) and \( g(e_2) = h(e_2) \). We claim that \( h = g \)—equivalently, \( g^{-1}h = \text{Id} \). Since \( g^{-1} \) is an affine transformation, our earlier observations show that \( f := g^{-1}h \) is a continuous function satisfying (*) and fixing \( 0, e_1 \), and \( e_2 \). We thus reduce to proving that such a function \( f \) must be the identity.

In the rest of the proof we assume that \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) satisfies (*). We first note that \( f \) must be injective: Indeed, suppose \( f(P) = f(Q) \), but \( P \neq Q \). Pick any point \( R \in \mathbb{R}^2 \) so that \( P, Q, R \) are not collinear. But \( f(R), f(P), f(Q) \) are certainly collinear, contradicting (*). We next note that if \( P, Q \) are distinct points of \( \mathbb{R}^2 \), then the image
f(PQ) of the line PQ must be contained in the line \( f(P)f(Q) \). Finally, we note that if PQ, RS are parallel lines, then the lines \( f(P)f(Q) \) and \( f(R)f(S) \) must be parallel.

From these simple remarks, we next deduce the key observation: If \( f \) fixes three vertices \( P, Q, R \) of a parallelogram with vertices \( P, Q, R, S \), then it must also fix the forth vertex \( S \). To see this, we can assume the vertices are labelled so that the lines \( PQ \) and \( RS \) are parallel and the lines \( PS \) and \( QR \) are parallel. Since \( f \) fixes \( P \) and \( Q \) and satisfies \((*)\), it takes the line \( PQ \) into itself, so, since \( f \) also fixes \( R \), it takes \( RS \) into \( Rf(S) \), hence, since \( f \) takes parallel lines to parallel lines, \( f(S) \) must lie on the line \( RS \). But we see similarly that \( f(S) \) must also lie on the line \( PS \), so \( f \) must fix \( S \).

Now, if \( f \) fixes \( 0 \), \( e_1 \), \( e_2 \), then we can repeatedly apply the above observation to parallelograms of one of the following five forms

1. \( P, P + e_1, P + e_2, P + e_1 + e_2 \)
2. \( P, P + e_1, P + e_1 + e_2, P + 2e_1 + e_2 \)
3. \( P, P + e_1, P + e_2, P - e_1 - e_2 \)
4. \( P, P + e_1, P + e_2, P + e_1 - e_2 \)
5. \( P, P + e_1, P - e_2, P + e_1 + e_2 \)

with \( P \in \mathbb{Z}^2 \subseteq \mathbb{R}^2 \) chosen appropriately to see that \( f \) must fix each point of \( \mathbb{Z}^2 \subseteq \mathbb{R}^2 \). For example: We first apply the observation to the parallelogram of the first type with \( P = 0 \) to see that \( f \) fixes \( e_1 + e_2 \). Next we repeatedly apply the observation to parallelograms of the second and third types to see that \( f \) fixes all points of the form \((a, b)\) with \( a \in \mathbb{Z} \), \( b \in \{0, 1\} \). Finally we apply the observation repeatedly to parallelograms of the fourth and fifth types to conclude that \( f \) fixes every point of \( \mathbb{Z}^2 \).

Next we argue that if \( f \) fixes \( 0 \), \( e_1 \), and \( e_2 \), then it must fix all points with rational coordinates. Consider a point \( S = (x, y) \in \mathbb{Q}^2 \subseteq \mathbb{R}^2 \). Take any point \( P \in \mathbb{Z}^2 \) not equal to \( S \). Since the line \( PS \) has rational slope and contains \( P \in \mathbb{Z}^2 \), it will contain some point \( R \in \mathbb{Z}^2 \) distinct from \( P \). Since \( f \) fixes \( P \) and \( R \) it takes \( PR \) into itself, hence \( f(S) \in PR = PS \). Now just take another point \( P' \in \mathbb{Z}^2 \) not lying on the line \( PR = PS \) and repeat the same argument to show that \( f(S) \in P'S \), hence \( f(S) = S \) as desired.

Now we have shown that a function \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) satisfying \((*)\) and fixing \( 0, e_1, e_2 \) fixes all points with rational coordinates, hence, if \( f \) is also continuous, it must be the identity. \(\square\)

**Remark 1.4.6.** For a general field \( \mathbb{K} \), a function \( f : \mathbb{K}^2 \to \mathbb{K}^2 \) satisfying \((*)\) in Theorem 1.4.5 is called a **collineation**. (One defines a line in \( \mathbb{K}^2 \) to be a subset of \( \mathbb{K}^2 \) arising as a translation of a one-dimensional linear subspace of the vector space \( \mathbb{K}^2 \).) Here is one way to construct collineations: Let \( G \) be the automorphism group of the field \( \mathbb{K} \) (the “Galois group” of \( \mathbb{K} \)). Each \( \sigma \in G \) gives a collineation \( f_{\sigma} : \mathbb{K}^2 \to \mathbb{K}^2 \) by applying \( \sigma \) to the coordinates; let us call such a collineation a **Galois collineation**. Notice that each Galois collineation fixes \((0, 0)\), \((1, 0)\), and \((0, 1)\). In fact, it can be shown that any collineation of \( \mathbb{K}^2 \) fixing \((0, 0)\), \((1, 0)\), and \((0, 1)\) is a Galois collineation. (The argument for this is elementary—I have omitted it only because it is a bit long and tedious.) It
follows from the argument in the above proof that any collineation of \( \mathbb{R}^2 \) can be written as a composition of an affine transformation of \( \mathbb{K}^2 \) and a Galois collineation. It can also be shown that the automorphism group of the field \( \mathbb{R} \) is trivial (the idea here is that the ordering of \( \mathbb{R} \) can be interpreted in terms of the field structure because, in \( \mathbb{R} \), being \( \geq 0 \) is the same thing as being a square, which is a field-theoretic notion). It follows that every collineation of \( \mathbb{R}^2 \) is, in fact, an affine transformation. In other words, Theorem 1.4.5 remains true even if the word “continuous” is deleted. See Propositions 3.11 and 3.12 in [H2].

**Remark 1.4.7.** The proof of Theorem 1.4.5 shows that any collineation \( f : \mathbb{Q}^2 \to \mathbb{Q}^2 \) is an affine transformation.

### 1.5 Classification of isometries of the plane

In this section we will classify all isometries of \( \mathbb{R}^2 \).

**Lemma 1.5.1.** Let \( C_1, C_2 \) be circles in \( \mathbb{R}^2 \) with centers \( P_1, P_2 \), respectively. Assume that \( C_1 \) and \( C_2 \) intersect in precisely two points: \( P \) and \( Q \). (In particular this implies that \( P_1 \neq P_2 \).) Then the line segment \( PQ \) is orthogonal to the line \( P_1P_2 \) and \( PQ \) intersects \( P_1P_2 \) at the midpoint of \( PQ \).

**Proof.** Let \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) be the isometry of \( \mathbb{R}^2 \) given by reflection (Example 1.3.6) across the line \( L = P_1P_2 \). Since \( P_1, P_2 \in L \), we have \( f(P_i) = P_i \) (\( i = 1, 2 \)). Since \( f \) is an isometry with \( f(P_i) = P_i \), \( f \) takes \( C_i \) into itself, hence \( f(P), f(Q) \) are in \( C_1 \cap C_2 = \{P, Q\} \). We cannot have \( f(P) = P \) and \( f(Q) = Q \), for then \( P, Q, \) and \( P_i \) would all lie on \( L \) (the fixed locus of \( f \)), and \( P, Q \) would be equidistant from both \( P_1 \) and \( P_2 \), hence we would have \( P = Q \). So, since \( f \) is bijective, we must have \( f(P) = Q \) and hence \( P = f(f(P)) = f(Q) \). The conclusion follows easily. \( \square \)

**Lemma 1.5.2.** Suppose \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is a distance preserving function such that \( f(P_1) = P, f(P_2) = P_2 \) and \( f(P_3) = P_3 \) for three non-collinear points \( P_1, P_2, P_3 \in \mathbb{R}^2 \). Then \( f \) is the identity.

**Proof.** We need to show that \( f(Q) = Q \) for an arbitrary \( Q \in \mathbb{R}^2 \). Suppose—toward a contradiction—that \( f(Q) \neq Q \) for some \( Q \in \mathbb{R}^2 \). Then \( Q \) cannot be one of the \( P_i \), so the distances \( d_i := d(P_i, Q) \) are positive. Let \( C_i \) be the circle of radius \( d_i \) centered at \( P_i \). Then \( Q \in C_1 \cap C_2 \cap C_3 \). Since \( f \) preserves distance and fixes \( P_i \) it takes \( C_i \) into \( C_i \), hence \( f(Q) \in C_1 \cap C_2 \cap C_3 \). Suppose \( i, j \in \{1, 2, 3\} \) are distinct. Then the circles \( C_i \cap C_j \) cannot intersect in more than two points, for then they’d be equal and we’d have \( P_i = P_j \), contradicting the assumption that the \( P_i \) aren’t collinear. We conclude that \( C_i \cap C_j = \{Q, f(Q)\} \). Let \( L_{ij} \) be the line through \( P_i \) and \( P_j \). By the previous lemma, the lines \( L_{12}, L_{13}, \) and \( L_{23} \) are all orthogonal to \( f(Q) \) and each intersects \( f(Q) \) at the midpoint of the line segment \( f(Q) \), so we have \( L_{12} = L_{13} = L_{23} \), contradicting the assumption that the \( P_i \) are not collinear. \( \square \)

**Lemma 1.5.3.** A function \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is distance preserving and satisfies \( f(0) = 0 \) iff \( f \) is an orthogonal linear transformation.
1.5 Classification of isometries of the plane

Proof. Since any linear transformation fixes zero (takes 0 to 0), Proposition 1.3.1 implies that any orthogonal linear transformation is an isometry fixing zero. Now suppose \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) is distance preserving and satisfies \( f(0) = 0 \). Then, since \( f \) preserves distance, \( f(e_1) \) and \( f(e_2) \) have to be on the unit circle (i.e. at distance one from the origin) and
\[
\| f(e_1), f(e_2) \| = \| e_1, e_2 \| = \sqrt{2}.
\]
As in the proof of Proposition 1.3.1, the Law Of Cosines then implies that \( f(e_1) \) and \( f(e_2) \) are orthogonal unit vectors, so the matrix \( A \) with columns \( f(e_1), f(e_2) \) is an orthogonal matrix and the corresponding linear transformation \( f_A : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) is an orthogonal linear transformation with \( f_A(e_1) = f(e_1), f_A(e_2) = f(e_2) \), and \( f_A(0) = 0 = f(A) \). Then the composition \( f_A^{-1}f \) is an isometry of \( \mathbb{R}^2 \) fixing the three non-collinear points 0, \( e_1, e_2 \), hence \( f_A^{-1} = \text{Id} \) by the previous lemma, hence \( f = f_A \). \( \square \)

Theorem 1.5.4. Every distance preserving function \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) can be written uniquely as the composition of an orthogonal linear transformation followed by a translation. In particular, every such \( f \) is an isometry.

Proof. Set \( t := -f(0) \in \mathbb{R}^2 \), and let \( f_t : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) be the corresponding translation. Then the composition \( f_t f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) is a distance preserving function satisfying \( f_t f(0) = 0 \), so by the previous lemma we have \( f_t f = g \) for some orthogonal linear transformation \( g : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \), and hence \( f = f_t g \) displays \( f \) as a composition of an orthogonal linear transformation followed by a translation. The uniqueness of this expression is established in the same manner one establishes uniqueness of the corresponding expression for affine transformations in §1.4. \( \square \)

Corollary 1.5.5. Every distance preserving function \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) is an affine transformation.

The following corollary could also be proved directly:

Corollary 1.5.6. Every isometry \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) can be expressed as a composition of reflections (Example 1.3.6), hence the group \( \text{Isom}(\mathbb{R}^2) \) can be generated by elements of order 2.

Proof. In light of the theorem it suffices to show that every translation and every orthogonal transformation can be expressed as a composition of reflections. Given two parallel lines \( L, L' \), the composition \( r_{L'} r_L \) of the corresponding reflections will be the translation \( f_{2t} \), where \( t \in \mathbb{R}^2 \) is the vector such that \( L' = f_t(L) \) (exercise!), so by choosing \( L \) and \( L' \) appropriately, we can realize every translation as a composition of (at most) two reflections.

In Example 1.3.6 we noted that every orthogonal translation of \( \mathbb{R}^2 \) is either a rotation around the origin, or a reflection over the \( x \)-axis followed by such a rotation. It thus remains only to prove that the counter-clockwise rotation \( f_\theta \) around the origin through an angle \( \theta \) can be expressed as a composition of reflections. Let \( L \) (resp. \( L' \)) be the line through the origin obtained by rotating the \( x \)-axis counter-clockwise around the origin through the angle \( \theta/2 \) (resp. \( \theta \)).
I claim that the composition \( r_L r_L' \) is \( f_\theta \). As discussed in Example 1.3.6, one can see by linear algebraic methods that both \( r_L \) and \( r_L' \) are orthogonal transformations of determinant \(-1\), hence \( r_L r_L' \) is an orthogonal transformation of determinant 1, so it must be a counter-clockwise rotation around the origin through \( \theta \) angle by the classification of orthogonal transformations in Proposition 1.3.3. This reduces us to showing that \((r_L r_L')(e_1)\) is equal to \( f_\theta(e_1) \), which is \((\cos \theta, \sin \theta)\). This is exactly what we arranged by our choice of \( L \) and \( L' \): We have \( r_L(e_1) = f_\theta(e_1) \) by the choice of \( L \), and this point is fixed by \( r_L' \) since it lies on \( L' \), by our choice of \( L' \).

**Remark 1.5.7.** The proof of the above corollary shows that every isometry \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is the composition of at most five reflections. In fact at most three are needed.

In light of the classification of isometries of \( \mathbb{R}^2 \) in Theorem 1.5.4, we can make the following definition, which will be useful later:

**Definition 1.5.8.** An isometry \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is called orientation preserving iff, in the unique expression \( f = f g \) of \( f \) as a composition of an orthogonal transformation \( g \) followed by a translation \( f_t \), the orthogonal transformation \( g \) is a special orthogonal linear transformation (i.e. a rotation—see Example 1.3.5).

**Remark 1.5.9.** It is possible to define the notion of “orientation preserving” for any homeomorphism \( f : \mathbb{R}^n \to \mathbb{R}^n \) in a manner restricting to the notion in Definition 1.5.8 for isometries of \( \mathbb{R}^2 \) and satisfying the expected properties: The identity map is orientation preserving and a composition \( f g \) of homeomorphisms is orientation preserving iff either both \( f \) and \( g \) are orientation preserving, or neither \( f \) nor \( g \) is orientation preserving.

### 1.6 Algebraic automorphisms

Let \( \mathbb{K} \) be a field. Let us agree that a function \( f : \mathbb{K}^n \to \mathbb{K}^m \) is algebraic iff there are polynomials

\[
f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n) \in \mathbb{K}[x_1, \ldots, x_m]
\]

such that

\[
f(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))
\]

for every \( x = (x_1, \ldots, x_n) \in \mathbb{K}^n \). A function \( f : \mathbb{K}^n \to \mathbb{K}^n \) is called an algebraic automorphism of \( \mathbb{K}^n \) iff \( f \) is bijective and both \( f \) and \( f^{-1} \) are algebraic.

For example, every affine transformation is an algebraic automorphism. One can check that every algebraic automorphism of \( \mathbb{K}^1 \) is of the form \( x \mapsto \lambda x + a \) for some \( \lambda \in \mathbb{K}^* \), \( a \in \mathbb{K} \)—in other words, every algebraic automorphism of \( \mathbb{K}^1 \) is an affine transformation. For \( \mathbb{K}^2 \), however, there are algebraic automorphisms \( \mathbb{K}^2 \to \mathbb{K}^2 \) which are not affine transformations:

**Example 1.6.1.** The function \( f : \mathbb{K}^2 \to \mathbb{K}^2 \) defined by \( f(x, y) := (x, x^2 + y) \) is an algebraic automorphism of \( \mathbb{K}^2 \) with inverse \( g \) given by \( g(x, y) = (x, x^2 - y) \). The reader can check as an exercise that \( f \) is not an affine transformation (unless \( \mathbb{K} = \mathbb{F}_2 \) is “the” field with two elements, for then \( x^2 = x \) for “every” (i.e. “both”) \( x \in \mathbb{F}_2 \)).
The Jacobian Conjecture is a famous unsolved conjecture about algebraic automorphisms of \( \mathbb{C}^n \). It says that if \( f_1, \ldots, f_n \in \mathbb{C}[x_1, \ldots, x_n] \) are polynomials such that the polynomial
\[
\det \begin{pmatrix}
\frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n}
\end{pmatrix}
\]
(the Jacobian determinant) is a non-zero constant, then the corresponding algebraic function \( f : \mathbb{C}^n \to \mathbb{C}^n \) is an algebraic automorphism. As mentioned above, this is not hard to see when \( n = 1 \), though it is unknown even when \( n = 2 \).

1.7 Exercises

Exercise 1.1. Let \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) be the linear transformation corresponding to the (invertible) matrix
\[
A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.
\]
Let \( S^1 := \{ x \in \mathbb{R}^2 : d(x, 0) = 1 \} \) be the unit circle. Draw a picture of the image \( f(S^1) \) of \( S^1 \) under \( f \) and a picture of \( f(f(S^1)) \).

Exercise 1.2. Let \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) be the linear transformation corresponding to the matrix
\[
A_\theta := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}
\]
(for some \( \theta \in \mathbb{R}/2\pi \mathbb{Z} \)). Show, by explicit calculation with the distance formula, that \( f \) is an isometry.

Exercise 1.3. Prove that the special orthogonal group \( \text{SO}_2 \) is a normal subgroup of the orthogonal group \( \text{O}_2 \).

Exercise 1.4. For points \( x, y \in \mathbb{R}^2 \), the line segment \( \overline{xy} \) is the set of points \( z \) in \( \mathbb{R}^2 \) that can be written in the form \( z = tx + (1 - t)y \) for some \( t \in [0, 1] \). Prove that a point \( z \in \mathbb{R}^2 \) lies on the line segment \( \overline{xy} \) iff the triangle inequality \( d(x, y) \leq d(x, z) + d(z, y) \) is actually an equality.

Exercise 1.5. Suppose \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is an isometry and let \( L \subseteq \mathbb{R}^2 \) be a line. Prove “directly” (without making use of the classification of isometries of \( \mathbb{R}^2 \)) that \( f(L) \subseteq \mathbb{R}^2 \) is a line. For clarity, let us take as the definition of a line: a subset \( L \subseteq \mathbb{R}^2 \) for which there exist real numbers \( a, b, c \) (with \( a, b \) not both equal to zero) such that
\[
L = \{ x = (x_1, x_2) \in \mathbb{R}^2 : ax_1 + bx_2 + c = 0 \}.
\]
(The previous exercise might be helpful.)
Chapter 2

Parametric Plane Curves

There are many possible things one can mean by a “curve” in the plane $\mathbb{R}^2$. In this chapter we will return to the study of parametrized plane curves, which should be familiar from calculus classes. Here are the main definitions:

**Definition 2.0.1.** A parametrized plane curve is a function $\gamma : [a, b] \to \mathbb{R}^2$, defined on some closed interval $[a, b] \subseteq \mathbb{R}$. We use $t$ to denote a point of $[a, b]$ and $x, y : [a, b] \to \mathbb{R}$ to denote the coordinates of $\gamma$, so that $\gamma(t) = (x(t), y(t))$. We assume that the derivatives $x', y', x'', y''$ with respect to $t$ exist and are continuous. We call $\gamma' := (x', y')$ the tangent vector to $\gamma$ and we say that $\gamma$ is smooth iff $\gamma'(t) \neq (0, 0)$ for all $t \in [a, b]$. A smooth parametrized plane curve $\gamma$ is said to be parameterized by arc length iff $a = 0$ and $|\gamma'(t)| = 1$ for all $t \in [0, b]$—in this case we generally use $s$ instead of $t$ for the parameter and we write $\ell$ instead of $b$, for reasons that will become clear momentarily.

**Example 2.0.2.** Fix a positive real number $r$. The function

\[
\gamma : [0, 2\pi] \to \mathbb{R}^2 \\
\gamma(t) := (r \cos t, r \sin t)
\]

is a smooth parametrized curve whose image $\gamma([0, 2\pi])$ is a circle of radius $r$ centered at the origin. The function

\[
\eta : [0, 2\pi r] \to \mathbb{R}^2 \\
\eta(s) := (r \cos(s/r), r \sin(s/r))
\]

is a smooth parametrized plane curve parametrized by arc length with the same image as $\gamma$.

In this chapter, we will use the word curve (resp. curve parameterized by arc length) to mean smooth parameterized plane curve (resp. smooth parameterized plane curve parameterized by arc length). We will study various geometric quantities associated with curves—for example, the length. Our assumption that the derivatives $x'$ and $y'$ exist and are continuous ensures that we can define the length in a fairly simple manner. While there are much more general definitions of length, they are less amenable to calculation.
and, in any case, are beyond the scope of these notes. The other important invariant attached to a curve is its curvature. Since our curves are plane curves, one can actually attach a sign to the curvature (this is not usually done in calculus classes) to define the signed curvature of a curve \( \gamma \) — this is a continuous function \( k : [a, b] \to \mathbb{R} \). This is ensured by our assumption that the second derivatives \( x'' \) and \( y'' \) exist and are continuous.

Here I want to focus on the interaction between the group of isometries of the plane and the properties of plane curves. The upshot is that the length and curvature are isometry invariant and they are, in some sense we will make precise, the “only” isometry invariants of a curve.

2.1 Reparametrization

The reader will probably agree that the two curves \( \gamma, \eta \) in Example 2.0.2 are in some sense “the same.” After all, the function \( \gamma \) is just the composition of multiplication by \( r \), viewed as a function \([0, 2\pi] \to [0, 2\pi r]\), followed by \( \eta \). In this section we are going to make this notion of “the same” precise. This point is implicit in the calculus treatment of parametrized curves, but it never really made explicit, probably because it is a bit technical.

Consider two closed intervals \([a, b], [c, d] \subseteq \mathbb{R}\). Let us use \( s \) for the coordinate on \([c, d]\) and \( t \) for the coordinate on \([a, b]\) for clarity. Let us agree that a nice bijection \( g : [c, d] \to [a, b] \) is a bijection \( g : [c, d] \to [a, b] \) which is twice continuously differentiable with \( g'(s) > 0 \) for all \( s \in [c, d] \).

It follows from the Implicit Function Theorem (see Theorem 6.3.1 in the appendix for a precise statement) that the inverse of \( g \) is also a nice bijection. Let us discuss this point a bit further, at it is rarely explained well in an introductory calculus sequence. For clarity, let us denote the inverse of \( g \) by \( f = f(t) : [a, b] \to [c, d] \), rather than by \( g^{-1} \). Then we have \( fg = \text{Id} \), or, in other words, \( f(g(s)) = s \) for every \( s \in [c, d] \). If we differentiate both sides of this equality with respect to \( s \) by using the Chain Rule, we find that

\[
  f'(g(s))g'(s) = 1. \tag{2.1}
\]

Then, using the fact that \( g'(s) > 0 \), we solve (2.1) for \( f'(g(s)) \) to find \( f'(g(s)) = 1/g'(s) \). Since every \( t \in [a, b] \) is \( g(s) \) for some (in fact a unique) \( s \in [c, d] \), we conclude that \( f'(t) > 0 \) for all \( t \in [a, b] \). Notice that we have to use the Implicit Function Theorem here to know a priori that \( f \) is differentiable, so that the Chain Rule can be applied to calculate the derivative of the composition \( f(g(s)) \). We can use the same method to calculate the second derivative of \( g \), since we know from the Implicit Function Theorem that it exists (and is continuous): We just differentiate (2.1) again with respect to \( s \) to find

\[
  f'(g(s))g''(s) + g'(s)^2 f''(g(s)) = 0
\]

and then we use that \( g'(s) > 0 \) to solve for \( f''(g(s)) \). This gives

\[
  f''(g(s)) = -\frac{f'(g(s))g''(s)}{g'(s)^2}.
\]
Chapter 2  Parametric Plane Curves

The Implicit Function Theorem says that this formula is indeed valid, so \( f'' \) exists and is continuous because \( g'' \) exists and is continuous.

It is customary in the situation described above to get rid of the notation for \( g \) and its inverse, and just view \( t = t(s) \) as a function of \( s \), and \( s = s(t) \) as a function of \( t \). With this understanding, formula (2.1) is rewritten

\[
\frac{ds}{dt} \frac{dt}{ds} = 1. \tag{2.2}
\]

**Definition 2.1.1.** If \( \gamma : [a, b] \to \mathbb{R}^2 \) is a (smooth parametrized plane) curve and \( t(s) : [c, d] \to [a, b] \) is a nice bijection, then the composition \( \gamma_t : [c, d] \to \mathbb{R}^2 \) is also a curve (by the Chain Rule). A curve obtained from \( \gamma \) in this manner is called a reparametrization of \( \gamma \). Two curves \( \gamma : [a, b] \to \mathbb{R}^2, \eta : [c, d] \to \mathbb{R}^2 \) are called equivalent iff \( \eta \) is a reparametrization of \( \gamma \). A parametric plane curve is an equivalence class of smooth parametrized plane curves.

The notion of “equivalence” defined above is indeed an equivalence relation on the set of curves. The observation at the beginning of this section implies that the two curves in Example 2.0.2 are equivalent.

### 2.2 Length and the arc length parametrization

Among all reparametrizations of a given curve, there is one particularly nice choice, called the arc length parametrization, which we will now describe.

Given a curve \( \gamma = (x, y) : [a, b] \to \mathbb{R}^2 \), recall that we call \( \gamma' = \gamma'(t) = (x'(t), y'(t)) \) the tangent vector to \( \gamma \) at \( \gamma(t) \). The smoothness assumption on \( \gamma \) says that \( \gamma' \) is never the zero vector, hence its magnitude

\[
|\gamma'| = \sqrt{(x')^2 + (y')^2}
\]

is a function \( |\gamma'| : [a, b] \to \mathbb{R}_{>0} \).

We define the length \( \ell = \ell(\gamma) \) of \( \gamma \) to be

\[
\ell(\gamma) := \int_a^b |\gamma'(t)| \, dt. \tag{2.3}
\]

More generally, we define the arc length function \( s : [a, b] \to [0, \ell] \) by

\[
s(t) := \int_a^t |\gamma'(\tau)| \, d\tau.
\]

By the Fundamental Theorem of Calculus we have

\[
\frac{ds}{dt} = |\gamma'(t)|, \tag{2.4}
\]

which is \( > 0 \) by the assumption that \( \gamma \) is smooth. Our assumption that \( x'' \) and \( y'' \) exist and are continuous also shows that

\[
\frac{d^2s}{dt^2} = \frac{x'x'' + y'y''}{|\gamma'(t)|}.
\]
exists and is continuous, hence our arc length function $s$ is a nice bijection, and hence so is its inverse

$$t = t(s) : [0, \ell] \to [a, b].$$

As discussed in §2.1, we have

$$\frac{dt}{ds} = \frac{1}{ds/dt} = \frac{1}{|\gamma'(s(t))|}.$$  \hspace{1cm} (2.5)

The reparametrization $\eta(s) := \gamma(t(s)) : [0, \ell] \to \mathbb{R}^2$ of $\gamma$ is called the arc length parametrization of $\gamma$. As the terminology suggests, the curve $\eta$ is parameterized by arc length because we compute

$$\left| \frac{d\eta}{ds} \right| = \left| \frac{d\gamma}{dt} \right| \left| \frac{dt}{ds} \right| = \left| \frac{d\gamma}{ds} \right| = 1$$

using the Chain Rule and (2.5).

The upshot of this section is that if one considers curves up to reparametrization, then nothing is lost (in theory) by considering only curves parametrized by arc length. However, it should be mentioned that the integrals defining $\ell(\gamma)$ and $s(t)$ can rarely be carried out in terms of elementary functions, so if one ever needs to explicitly compute a quantity associated to the arc length parametrization of a curve $\gamma$, one usually tries to express that quantity in terms of the original parametrization of $\gamma$ by using the Chain Rule.

### 2.3 Curvature and the Frenet formulas

Consider a curve $\gamma = \gamma(s) : [0, \ell] \to \mathbb{R}^2$ parametrized by arc length. We call $T(s) := \gamma'(s)$ the unit tangent vector to $\gamma$ at $\gamma(s)$. Since $\gamma$ is parametrized by arc length, we have $|T| = 1$, independent of $s$. By linear algebra, there is a unique unit vector $N = N(s)$ orthogonal to $T$ such that the matrix $(T, N)$ with columns given by $T$ and $N$ has determinant 1. The vector $N(s)$ is called the principal unit normal to $\gamma$ at $\gamma(s)$. Since $|T| = 1$ is constant, if we differentiate both sides with respect to $s$ we find that $T \cdot T' = 0$. So $T'$ is orthogonal to $T$, hence it must be some scalar multiple of $N$:

$$T' = kN.$$  \hspace{1cm} (2.6)

The scalar multiple $k = k(s)$ is called the signed curvature of $\gamma$ at $\gamma(s)$. The formula for $k$ in Exercise 2.3 makes it clear that $k$ is a continuous function $k : [0, \ell] \to \mathbb{R}$. 

One can now consider the derivative $N'$. By the same arguments used for $T$, we see that $N'$ will be some scalar multiple of $T$: $N' = \alpha T$. In fact this $\alpha$ doesn’t give us anything “new”: By differentiating $(N, T) = 0$, we find that

\[ (N', T) + (N, T') = 0. \]

Using the definitions of the signed curvature $k$ and $\alpha$ we can rewrite this

\[ (\alpha T, T) + (N, kN) = 0. \]

Using bilinearity of the inner product and the fact that $T$ and $N$ are unit vectors, this gives $\alpha + k = 0$, or $\alpha = -k$. The discussion thus far establishes the Frenet formulas (for curves in the plane parametrized by arc length):

\[ T' = kN \]
\[ N' = -kT. \]  

**Example 2.3.1.** Consider the curve parametrized by arc length $\eta : [0, 2\pi r] \to \mathbb{R}^2$ from Example 2.0.2 whose image is the circle of radius $r$ centered at the origin in $\mathbb{R}^2$. Note that $\eta$ traces out its image in the counter-clockwise direction. We have

\[ T = (r \cos(s/r), r \sin(s/r))' \]
\[ = (-\sin(s/r), \cos(s/r)). \]

Since

\[
\begin{pmatrix}
- \sin(s/r) - \cos(s/r) \\
\cos(s/r) - \sin(s/r)
\end{pmatrix}
\]

is an orthogonal matrix, we have

\[ N = (- \cos(s/r), - \sin(s/r)). \]

We compute

\[ T' = (-1/r) \cos(s/r), -(1/r) \sin(s/r)) \]
\[ = (1/r)N, \]

so the signed curvature $k$ of $\eta$ is constant, equal to $1/r$.

If there is any chance of confusion (for example, if we consider two curves parametrized by arc length at the same time) then we will write $T_\gamma$, $N_\gamma$, and $k_\gamma$ instead of $T$, $N$, and $k$.

As mentioned in the previous section, the arc length parametrization of a curve $\gamma = \gamma(t)$ is generally difficult, or—in some sense—impossible, to write down, so it is worth having formulas for $T$, $N$, and $k$ in terms of an arbitrary parametrization. This is done by using formula (2.5), together with the Chain Rule, to re-express the derivatives.
with respect to arc length \( s \) used to define \( T, N, \) and \( k \) in terms of differentiation with respect to \( t \). For example,

\[
T(t) = \frac{d\gamma(t(s))}{ds} = \frac{d\gamma}{dt} \frac{dt}{ds} = \frac{1}{|\gamma'(t)|} \gamma'(t),
\]

where the primes denote differentiation with respect to \( t \). Then \( N(t) \) is defined from \( T(t) \) in the same way that \( N(s) \) is defined from \( T(s) \). Similarly we compute

\[
\frac{dT}{ds} = \frac{d\gamma}{dt} \frac{d\gamma}{ds} = \frac{1}{|\gamma'|} T',
\]

where the primes denote derivatives with respect to \( t \). Then \( k(t) \) is defined by

\[
\frac{dT}{ds} = k(t)N(t).
\]

2.4 Plane curves up to isometry

If \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is any nice enough function (a smooth bijection with smooth inverse, say), and \( \gamma : [a, b] \to \mathbb{R}^2 \) is a smooth parametrized curve, then the composition \( f\gamma : [a, b] \to \mathbb{R}^2 \) is also a smooth parametrized curve. We will be interested in this construction only when \( f \) is an orientation preserving isometry (Definition 1.5.8) and \( \gamma : [0, \ell] \to \mathbb{R}^2 \) is a curve parametrized by arc length. In this situation we can write \( f = fg \), where \( g \in SO_2 \) and \( f_1 \) is a translation. Let us write \( A = A_g \) for the orthonormal matrix corresponding to \( g \). One checks (Exercise 2.4) that:

1. \((f\gamma)' = A\gamma'\), hence \( f\gamma \) is also parameterized by arc length and \( T_{f\gamma} = AT_{\gamma} \).
2. \( \ell(f\gamma) = \ell(\gamma) \).
3. The principal unit normal vectors of \( \gamma \) and \( f\gamma \) are related by \( N_{f\gamma} = AN_{\gamma} \).
4. The curvatures of \( \gamma \) and \( f\gamma \) are equal: \( k_\gamma = k_{f\gamma} \).

**Theorem 2.4.1.** Suppose \( \gamma : [0, \ell] \to \mathbb{R}^2 \) and \( \overline{\gamma} : [0, \overline{\ell}] \to \mathbb{R}^2 \) are two curves parametrized by arc length. Then the following are equivalent:

1. \( \gamma = f\gamma \) for some orientation preserving isometry \( f : \mathbb{R}^2 \to \mathbb{R}^2 \).
2. The curves \( \gamma \) and \( \overline{\gamma} \) have the same length and curvature: That is \( \ell = \overline{\ell} \) and \( k = \overline{k} \).
Proof. The fact that (1) implies (2) is discussed above and left for the reader as Exercise 2.4. To see that (2) implies (1), suppose \( \gamma \) and \( \tau \) are curves parametrized by arc length with the same length and curvature. Let \( t := \gamma(0) - \tau(0) \in \mathbb{R}^2 \) and let \( A \in SO_2 \) be the rotation taking the unit vector \( \tau'(0) \) to the unit vector \( \gamma'(0) \). Then \( f := f_t f_A \) is an isometry of \( \mathbb{R}^2 \) such that the curves \( \gamma \) and \( f_\tau \) have the same initial point and initial tangent vector. By the implication already “proved,” these two curves also have the same length and curvature. By replacing \( \tau \) with \( f_\tau \), we thus reduce to proving that if \( \gamma \) and \( \tau \) are curves parametrized by arc length with the same initial point, the same initial tangent vector, and the same length \( \ell \) and curvature \( k \), then they are equal. Let \( T \) and \( N \) (resp. \( \overline{T} \) and \( \overline{N} \)) be, respectively, the unit tangent vector and principal unit normal for the curve \( \gamma \) (resp. \( \tau \)). Using the Frenet Formulas (2.7) we compute

\[
\begin{align*}
\left( \langle T - \overline{T}, T - \overline{T} \rangle + \langle N - \overline{N}, N - \overline{N} \rangle \right)' \\
&= 2\langle T' - \overline{T}', T - \overline{T} \rangle + 2\langle N' - \overline{N}', N - \overline{N} \rangle \\
&= 2k\langle N - \overline{N}, T - \overline{T} \rangle - 2\langle kT - k\overline{T}, N - \overline{N} \rangle \\
&= -2k\langle N, \overline{T} \rangle - 2k\langle \overline{N}, T \rangle + 2k\langle N, \overline{T} \rangle + 2k\langle \overline{N}, T \rangle \\
&= 0.
\end{align*}
\]

This shows that the function in parentheses is constant—but it is zero at \( s = 0 \) since \( T(0) = \overline{T}(0) \) (hence \( N(0) = \overline{N}(0) \) as well), so it is zero. But this function is the sum of the magnitude of \( T - \overline{T} \) and the magnitude of \( N - \overline{N} \), both of which are non-negative, so we must have \( T = \overline{T} \). Since \( T = \gamma' \) and \( \overline{T} = \overline{\gamma}' \), this proves that the derivatives of the coordinate functions of \( \gamma \) and \( \overline{\gamma} \) are equal. But we also have \( \gamma(0) = \overline{\gamma}(0) \), hence \( \gamma = \overline{\gamma} \) by the Fundamental Theorem of Calculus. \( \square \)

Remark 2.4.2. There is an analogue of Theorem 2.4.1 for curves in \( \mathbb{R}^3 \) parametrized by arc length. In this case, the curvature cannot be given a sign, and one must also consider another invariant called the torsion. (The proof is essentially the same, though the Frenet formulas become a bit more complicated in three dimensions.) For this, see the first chapter of the book Differential Geometry of Curves and Surfaces by Manfredo P. Do Carmo, from which the material in this section has been “borrowed.”

2.5 Exercises

Exercise 2.1. Suppose \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is an isometry of \( \mathbb{R}^2 \) which is orientation reversing (not orientation preserving) and \( \gamma \) is a curve parametrized by arc length in \( \mathbb{R}^2 \). Show that \( f \gamma \) is a curve parametrized by arc length in \( \mathbb{R}^2 \) with signed curvature \( k_{f\gamma} = -k_\gamma \).

Exercise 2.2. Let \( P \) and \( Q \) be distinct points of \( \mathbb{R}^2 \). Let \( v := (Q - P)/|Q - P| \) be the unit vector pointing from \( P \) to \( Q \) and let \( \ell := |Q - P| \) be the distance from \( P \) to \( Q \). Show that the curve \( \gamma : [0, \ell] \to \mathbb{R}^2 \) defined by \( \gamma(s) := P + sv \) is a smooth curve parametrized by arc length whose signed curvature \( k \) is identically zero.

Exercise 2.3. Consider a smooth plane curve \( \gamma = \gamma(s) = (x(s), y(s)) : [0, \ell] \to \mathbb{R}^2 \) parametrized by arc length. Show that the signed curvature \( k \) of \( \gamma \) is given by

\[
k = x'y'' - x''y'.
\]
Hint: First figure out an explicit formula for the principal unit normal \( N \) in terms of \( x, y \) and their derivatives. You also have an obvious such formula for \( T' \), so you just have to check that the proposed formula for \( k \) actually satisfies \( T' = kN \). For this, it might be helpful to have the identity obtained by writing out the condition \( |T| = 1 \) in terms of \( x, y \) and their derivatives then differentiating both sides.

**Exercise 2.4.** Check the four listed assertions at the beginning of §2.4.
Chapter 3

Affine Algebraic Plane Curves

Generally speaking, a curve should be a subset of the plane $\mathbb{R}^2$ obtained by imposing some “reasonable” relationship between the two coordinates $x, y$ of a point $(x, y) \in \mathbb{R}^2$. For example, the graph $\Gamma_f \subseteq \mathbb{R}^2$ of a function $f: \mathbb{R} \to \mathbb{R}$ is a curve since it is obtained by requiring that the second coordinate be the value of $f$ on the first coordinate:

$$\Gamma_f := \{(x, y) \in \mathbb{R}^2 : y = f(x)\}.$$

It seems reasonable then, to say that an algebraic curve (more precisely, an affine algebraic plane curve) is a subset of $\mathbb{R}^2$ obtained by imposing an “algebraic” (i.e. polynomial) relationship between the coordinates. For example, if

$$f(x) = a_0 + a_1x + \cdots + a_nx^n$$

is a polynomial function of $x$, then the graph $\Gamma_f \subseteq \mathbb{R}^2$ is an algebraic curve. Here are some other algebraic curves that you should have encountered at some point:

$$C_1 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$$
$$C_2 := \{(x, y) \in \mathbb{R}^2 : x = 0\}$$
$$C_3 := \{(x, y) \in \mathbb{R}^2 : x^2 - y^2 = 1\}$$

The curve $C_1$ is a circle of radius one centered at the origin, $C_2$ is the $y$-axis, and $C_3$ is a hyperbola. Of course, not all plane curves are graphs of functions (e.g. $C_1$, $C_2$, $C_3$ above).

The notion of an algebraic curve makes sense over any field $K$. For example, replacing $\mathbb{R}$ everywhere with the complex numbers $\mathbb{C}$, one can also consider complex algebraic curves $C \subseteq \mathbb{C}^2$ in the complex plane $\mathbb{C}^2$. It may seem at first that this might complicate matters: For one thing, it becomes difficult to visualize a complex plane curve, since the complex plane $\mathbb{C}^2$ itself has “four real dimensions.” On the other hand, the field of complex numbers $\mathbb{C}$ has certain technical advantages over $\mathbb{R}$, particularly the fact that it is algebraically closed: Any polynomial (of positive degree in one variable) with complex coefficients can be factored as a product of linear (degree one) polynomials. In general, the theory of algebraic curves, as we have defined them, is most satisfactory over an algebraically closed field, where the relationship between a polynomial and the algebraic curve it defines can be most satisfactorily understood.
As this course is meant to be a course on projective geometry, we will not dwell overly on the subject of affine curves. Our purpose in this section is mainly just to give the reader a taste for the theory of algebraic plane curves, without with our later study of projective plane curves would seem rather unmotivated. In order for this chapter to have some actual content, I have chosen to discuss the problem of classifying affine conics (over \(\mathbb{R}\) and \(\mathbb{C}\)) up to an appropriate notion of affine equivalence. The classification will be given without proof, though we shall give a complete treatment of the analogous classification problem in the projective setting. Already in this classification problem we will see that \(\mathbb{C}\) has an important “advantage” over \(\mathbb{R}\): Every complex number is a square, whereas real numbers less than zero are not squares.

3.1 Polynomials and zero loci

Recall that a polynomial in a single variable \(x\) with coefficients in a field \(K\) is a formal expression \(f(x)\) of the form

\[
f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,
\]

where the coefficients \(a_i\) are elements of \(K\). The set of polynomials with coefficients in \(K\) is denoted \(K[x]\). This set carries various additional structure: It is an \(K\)-algebra, which basically means that one can add, subtract, and multiply polynomials in a meaningful way, and that one can regard an element of \(K\) as a constant polynomial.

Though it is perhaps a stretch to emphasize this point now, it might be worth explaining—for the more mathematically sophisticated student—what is meant by “formal expression.” We really mean here that a polynomial is just a way of writing (and thinking about) a sequence 

\[
(a_0, a_1, a_2, \ldots)
\]

of elements of \(K\) which is eventually zero. We can add, subtract, and multiply such sequences (by thinking of them as coefficients of polynomials). Now, of course, we can also view the polynomial as a function \(f: K \to K\), as we often do in calculus when \(K = \mathbb{R}\). It might not seem like such a big deal to distinguish between the formal list of coefficients and the corresponding function, and, indeed, it is not in many senses: you can recover the polynomial (i.e. its list of coefficients) from the corresponding function \(f : \mathbb{R} \to \mathbb{R}\) by the rule

\[
a_n = \frac{f^{(n)}(0)}{n!}
\]

(as you learn in the study of Taylor series). On the other hand, if you work with coefficients in a finite field \(K\), say, then of course we can still view a polynomial \(f(x) \in K[x]\) as a function \(f : K \to K\), but obviously there can be no hope of recovering the polynomial from the corresponding function because \(K[x]\) is infinite, while there are only finitely many functions \(K \to K\)! For example, if \(K = F_2 = \mathbb{Z}/2\mathbb{Z} = \{0, 1\}\) is the two-element field, then \(x = x^2\) as functions \(K \to K\), but we do not think of \(x = 0 + 1x\) and \(x^2 = 0 + 0x + 1x^2\) as being the same polynomial. This distinction is largely a pedagogical matter, but it might be wise for the reader to forget everything he/she knows about functions \(\mathbb{R} \to \mathbb{R}\) for a while and just think about polynomials in a completely formal way.
Definition 3.1.1. The degree of a polynomial $f$ as above is the largest $n$ for which the coefficient of $x^n$ in $f$ is non-zero:

$$\deg(f) := \max\{i : a_i \neq 0\}. $$

Thus a non-zero constant polynomial $f(x) = a_0 \neq 0$ has degree zero. It is often convenient to define the degree of the zero polynomial to be $-\infty$. The terms “non-constant” and “of positive degree” are equivalent; similarly “linear” and “degree one” are interchangeable.

We can make the same definitions for polynomials in two (or more) variables $x, y$. Now a polynomial is a formal expression of the form

$$f(x, y) = \sum_{i,j=0}^{\infty} a_{ij} x^i y^j,$$

where all but finitely many of the real numbers $a_{ij}$ are zero. The degree of such a polynomial is defined by

$$\deg(f) := \max\{i+j : a_{ij} \neq 0\}.$$

Evidently, any polynomial $f(x, y)$ of degree $d$ can be written

$$f = \sum_{i+j \leq d} a_{ij} x^i y^j$$

with at least one of $a_{d,0}, a_{d-1,1}, \ldots, a_{1,d-1}, a_{0,d}$ non-zero. The set $\mathbb{K}[x, y]$ of polynomials in variables $x, y$ forms an $\mathbb{K}$-algebra.

Definition 3.1.2. Given a polynomial $f(x, y) \in \mathbb{K}[x, y]$, we call the subset

$$Z(f) := \{(x, y) \in \mathbb{K}^2 : f(x, y) = 0\}$$

of $\mathbb{K}^2$ the zero locus of $f(x, y)$. A subset of $\mathbb{K}^2$ of the form $Z(f)$, for some non-constant $f(x, y) \in \mathbb{R}[x, y]$ will be called an affine algebraic plane curve over $\mathbb{K}$, or simply, an algebraic curve. If $C \subseteq \mathbb{K}^2$ is an algebraic curve, then any $f \in \mathbb{K}[x, y]$ for which $C = Z(f)$ will be called a defining equation for $C$.

When discussing algebraic curves, we often call $\mathbb{K}^2$ the affine plane over $\mathbb{K}$. Usually we refer to $\mathbb{R}^2$ simply as the plane and to $\mathbb{C}^2$ as the complex plane.

Remark 3.1.3. If the field $\mathbb{K}$ is contained in a larger field $\overline{\mathbb{K}}$, then of course we can also consider the zero locus of a polynomial $f(x, y) \in \mathbb{K}[x, y]$ in the larger affine plane $\overline{\mathbb{K}}^2$. If there is any chance of confusion, we will add some notation to $Z(f)$ to emphasize which field we are working with. We will write $Z(f)(\mathbb{K})$ for the zero locus of $f$ in $\mathbb{K}^2$, to distinguish it from the zero locus of $f$ in $\overline{\mathbb{K}}^2$, which we will denote $Z(f)(\overline{\mathbb{K}})$. The interested student may note that the Galois group $G = \text{Gal}(\overline{\mathbb{K}}/\mathbb{K})$ (the group of automorphisms of the field $\overline{\mathbb{K}}$ which restrict to the identity on the subfield $\mathbb{K} \subseteq \overline{\mathbb{K}}$) acts on $Z(f)(\overline{\mathbb{K}})$. The
student who wants to think about this at all should think about the case where the field extension is $\mathbb{R} \subseteq \mathbb{C}$, in which case the Galois group is the group with two elements because complex conjugation $z \mapsto \overline{z}$ is the unique non-identity automorphism of $\mathbb{C}$ restricting to the identity on $\mathbb{R}$. If $f \in \mathbb{R}[x, y]$ and $(a, b) \in Z(f)(\mathbb{C}) \subseteq \mathbb{C}^2$, then $(\pi, \overline{b}) \in Z(f)(\mathbb{C}) \subseteq \mathbb{C}^2$ as well. This is because, for any $a, b \in \mathbb{C}$, we have $\overline{f(a, \overline{b})} = f(\pi, \overline{b})$ since the coefficients of $f$ are real. It was realized early on in the development of algebraic geometry that the best “geometric” picture of the polynomial $f$ is obtained by considering all the sets $Z(f)(\overline{\mathbb{R}})$, for all extension fields $\mathbb{K} \subseteq \overline{\mathbb{R}}$, together with the actions of the relevant Galois groups. In modern algebraic geometry this information is all packaged together by considering the ring $\mathbb{K}[x, y]/(f)$ and its associated “prime spectrum.” The historical development of this perspective is described in [D].

For example, the subsets $C_1, C_2, C_3$ of $\mathbb{R}^2$ from the introduction to this chapter can be written $Z(f_1), Z(f_2), Z(f_3)$ for $f_1(x, y) = x^2 + y^2 - 1$, $f_2(x, y) = x$, $f_3(x, y) = x^2 - y^2 - 1$.

Since a product of two elements of $\mathbb{K}$ is zero iff at least one of them is zero, we have

$$Z(fg) = Z(f) \cup Z(g).$$

For example, $Z(xy)$ is the union of the $y$-axis and the $x$-axis.

The subset $Z(f)$ does not determine the polynomial $f$: It is certainly possible that $Z(f) = Z(g)$ for different polynomials $f, g$. For example, if $\lambda \in \mathbb{K}^* = \mathbb{K} \setminus \{0\}$, then clearly $Z(f) = Z(\lambda f)$ since $f(x, y) = 0$ iff $\lambda f(x, y) = 0$. For similar reasons,

$$Z(f) = Z(f^2) = Z(f^3) = \cdots.$$  

“Even worse,” there are lots of polynomials $f \in \mathbb{R}[x, y]$, such as $f(x, y) = x^2 + y^2 + 1$, for which $Z(f) \subseteq \mathbb{R}^2$ is empty, and lots of polynomials such as $f(x, y) = x^2 + y^2$, or $f(x, y) = x^4 + y^4$, for which $Z(f)$ consists of a single point.

At this point one might wonder whether “algebraic curve” is really a good name for the sets $Z(f)$ which can be finite (or empty), despite being called “curves,” which suggests they are one dimensional in some sense. One such sense is that the Krull dimension of the ring $\mathbb{K}[x, y]/(f)$ is 1 for every non-constant $f(x, y) \in \mathbb{K}[x, y]$ (by Krull’s Hauptidealzsatz). Furthermore, the examples where $Z(f)$ is empty or a finite set of points occur only when working over the real numbers $\mathbb{R}$. Over C, it turns out that, after possibly removing finitely many points from $Z(f)$, the set $Z(f)$ carries, in some reasonably natural way, the structure of a (non-empty!) complex one dimensional manifold. We will not use these facts in anything that follows; they are mentioned only to justify our terminology.

Over the complex numbers, one can, roughly speaking, recover $f$ from the subset $Z(f) \subseteq \mathbb{C}^2$ up to the sort of ambiguities discussed two paragraphs above. The precise statement is Hilbert’s Nullstellensatz which says that the subset (ideal, in fact)

$$\{g(x, y) \in \mathbb{C}[x, y] : g(x, y) = 0 \text{ for all } (x, y) \in Z(f) \subseteq \mathbb{C}^2\}$$

of $\mathbb{C}[x, y]$ coincides with the subset (ideal)

$$\{g(x, y) : g^n = hf \text{ for some } h \in \mathbb{C}[x, y] \text{ and some } n \in \{1, 2, \ldots \}\}.$$
We mention this only for the sake of completeness; we will not use this fact elsewhere.

Since the subset $\mathbb{Z}(f)$ is determined from $f$, but not vice-versa, we often think of the polynomial $f$ as being the more fundamental object. This point of view will be especially important in §3.3 where we discuss the problem of “classifying” the algebraic plane curves (of, say, a given degree).

In particular, we would like to say that the degree of an algebraic plane curve $C \subseteq \mathbb{K}^2$ is the degree of “the” polynomial $f$ for which $C = \mathbb{Z}(f)$. Of course this makes no sense because there are many such polynomials $f$! There are various ways of resolving this—we could, for example, consider the minimum degree of all those $f$ for which $C = \mathbb{Z}(f)$.

Rather than worry about this technicality, let us just agree to the following:

**Definition 3.1.4.** A line (resp. conic, cubic, quartic, . . . ) is an algebraic curve $C \subseteq \mathbb{K}^2$ which can be written $C = \mathbb{Z}(f)$ for some polynomial $f \in \mathbb{K}[x,y]$ of degree one (resp. two, three, four, . . . ).

Notice that, according to this definition, every “line” is also a conic, because if $C = \mathbb{Z}(f)$ with $f$ linear, then we also have $C = \mathbb{Z}(f^2)$, and $f^2$ will have degree two.

### 3.2 Smooth and singular points

Just as there is a notion of “smoothness” for parametric curves (Chapter 2), there is a very closely related notion of smoothness for algebraic curves.

Notice that the partial derivatives $f_x$ and $f_y$ of a polynomial $f \in \mathbb{K}[x,y]$ make sense over any field $\mathbb{K}$ (not just over the real numbers, where they have a “calculus” interpretation in terms of limits). Explicitly, if

$$f = \sum_{i,j} a_{i,j} x^i y^j$$

then we define $f_x$ and $f_y$ to be

$$f_x := \sum_{i,j} ia_{i,j} x^{i-1} y^j$$
$$f_y := \sum_{i,j} ja_{i,j} x^i y^{j-1}.$$ 

It is perhaps not so obvious that these “definitions” of the derivative should have any geometric significance over other fields, though we shall see that this is the case.

**Definition 3.2.1.** For $f \in \mathbb{K}[x,y]$, a point $P = (x_0, y_0) \in \mathbb{Z}(f) \subseteq \mathbb{K}^2$ is called a non-singular (or smooth) point of $\mathbb{Z}(f)$ iff at least one of $f_x(P), f_y(P)$ is non-zero. A point $P \in \mathbb{Z}(f)$ for which $f_x(P) = f_y(P) = 0$ is called a singular point of $\mathbb{Z}(f)$. The subset of $\mathbb{Z}(f)$ consisting of smooth (resp. singular) points is denoted $\mathbb{Z}(f)^{sm}$ (resp. $\mathbb{Z}(f)^{sing}$). We say that $f$ is non-singular or smooth iff $\mathbb{Z}(f)^{sing}(\mathbb{K})$ is empty for every field extension $\mathbb{K} \subseteq \mathbb{K}$.\(^1\)

\(^1\)It equivalent to ask that $\mathbb{Z}(f)^{sing}(\mathbb{K}) = \emptyset$ for some algebraically closed field $\mathbb{K}$ containing $\mathbb{K}$.
It is worth emphasizing that the question of whether \( P \in \mathbb{Z}(f) \) is a smooth point or a singular point depends highly on \( f \), not just on the curve \( C = \mathbb{Z}(f) \). For example, for any (non-constant) \( f \in \mathbb{K}[x, y] \), we have \( \mathbb{Z}(f) = \mathbb{Z}(f^2) \), but every point of \( \mathbb{Z}(f^2) \) is a singular point (Exercise 3.2).

In defining smoothness, we require that \( \mathbb{Z}(f)^{sing}(\mathbb{K}) \) be empty for all field extensions \( \mathbb{K} \supseteq \mathbb{K}(x, y) \), rather than just \( \mathbb{K} \) itself for various reasons. For one thing, if we have a polynomial \( f \in \mathbb{K}[x, y] \) and we regard it as a polynomial in \( f \in \mathbb{K}(x, y) \), we don’t want the meaning of “smooth” to change! A good example to keep in mind is the polynomial \( f = (1 + x^2)^2 \in \mathbb{R}[x, y] \), which has empty zero locus over \( \mathbb{R} \), even though every point of \( \mathbb{Z}(f^2) \) is singular (c.f. Exercise 3.2). This phenomenon illustrates the general philosophy of Remark 3.1.3 that one should consider all zero loci of \( f \) over all field extensions of \( \mathbb{K} \) together.

**Definition 3.2.2.** If \( P = (x_0, y_0) \) is a non-singular point of \( \mathbb{Z}(f) \), we define the tangent line to \( \mathbb{Z}(f) \) at \( P \) to be the line

\[
T_P \mathbb{Z}(f) := \mathbb{Z}(xf_x(P) + yf_y(P) - x_0f_x(P) - y_0f_y(P)).
\]

Notice that the assumption that \( P \) is non-singular is needed to ensure that the linear polynomial whose zero locus defines \( T_P \mathbb{Z}(f) \) is non-constant, so that \( T_P \mathbb{Z}(f) \) is actually a line. Note also that the constant term of the linear polynomial is chosen so that \( P \in T_P \mathbb{Z}(f) \).

**Proposition 3.2.3.** Let \( P = (a, b) \) be a smooth point of a curve \( \mathbb{Z}(f) \subseteq \mathbb{R}^2 \). Then there are open subsets \( U, V \) of \( \mathbb{R} \), containing \( a \) and \( b \), respectively, such that \( \mathbb{Z}(f) \cap (U \times V) \) is either the graph of a smooth (infinitely differentiable) function \( g : U \rightarrow V \) (this will be the case if \( f_x(P) \neq 0 \)) or the graph of a smooth function \( h : V \rightarrow U \) (if \( f_y(P) \neq 0 \)). In other words, at least one of the projections

\[
\pi_1 : \mathbb{Z}(f) \cap (U \times V) \rightarrow U
\]

\[
\pi_2 : \mathbb{Z}(f) \cap (U \times V) \rightarrow V
\]

will be a continuous bijection with smooth inverse.

**Proof.** This is a special case of the Implicit Function Theorem (Theorem 6.3.1 in the appendix).

### 3.3 Change of variables and affine equivalence

Recall the group \( \text{Aff}(\mathbb{K}^2) \) of affine transformations of the plane from §1.4. Given \([A, t] \in \text{Aff}(\mathbb{K}^2) \) and \( f \in \mathbb{K}[x, y] \), we define a new polynomial \( f \cdot [A, t] \in \mathbb{K}[x, y] \) by setting

\[
(f \cdot [A, t])(x, y) := f([A, t](x, y)).
\]

To be careful here, we should stick more precisely to the conventions of linear algebra and write every element of \( \mathbb{K}^2 \), as well as the argument of \( f \), as a column vector. Then
we would write:

\[(f \cdot [A, t]) \begin{pmatrix} x \\ y \end{pmatrix} := f(A \begin{pmatrix} x \\ y \end{pmatrix} + t).\]

To be completely explicit, if

\[A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{and} \quad t = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix},\]

then

\[f(x, y) = f(ax + by + t_1, cx + dy + t_2).\]

Usually I will be more sloppy with the notation since it is painful to write column vectors in the inline text. A good compromise is to write \(x\) (or \(x\) if you prefer) for the vector of variables \((x, y)\) (thought of as a column vector, but often abusively written as a row vector). Then \(f \cdot [A, t]\) is defined by

\[(f \cdot [A, t])(x) := f(Ax + t), \quad (3.1)\]

where the argument of \(f\) is thought of as a column vector.

The above construction of \(f \cdot [A, t]\) defines an action (more precisely, a right action) of \(\text{Aff}(\mathbb{K}^2)\) on \(\mathbb{K}[x, y]\). This means that

1. \(f \cdot \text{Id} = f\), where \(\text{Id} = [\text{Id}, 0]\) is the identity element of \(\text{Aff}(\mathbb{K}^2)\), and

2. \((f \cdot [A, t]) \cdot [A', t'] = f \cdot ([A, t][A', t'])\)

for every \(f \in \mathbb{K}[x, y]\), \([A, t], [A', t'] \in \text{Aff}(\mathbb{K}^2)\).

The first equality is obvious and the second is a simple calculation using definition (3.1) and the formula (1.3) for products in \(\text{Aff}(\mathbb{K}^2)\):

\[
((f \cdot [A, t]) \cdot [A', t'])(x) = (f \cdot [A, t])(A'x + t')
\]

\[
= f(A(A'x + t') + t)
\]

\[
= f(AA'x + At' + t)
\]

\[
= (f \cdot [AA', At' + t])(x)
\]

\[
= (f \cdot ([A, t][A', t']))(x).
\]

One can also think of this action by noting that any \([A, t] \in \text{Aff}(\mathbb{K}^2)\) gives rise to an automorphism of the \(\mathbb{K}\)-algebra \(\mathbb{K}[x, y]\) given by

\[x \mapsto ax + by + t_1\]

\[y \mapsto cx + dy + t_2,\]

where we have written out \(A\) and \(t\) as above.

More, general, the same formula (3.1) may be viewed as defining an action of \(\text{Aff}(\mathbb{K}^n)\)

on \(\mathbb{K}[x_1, \ldots, x_n]\) (think of \(x\) as the column vector of variables \((x_1, \ldots, x_n)\)).

The actions of \(\text{Aff}(\mathbb{K}^2)\) on \(\mathbb{K}^2\) and \(\mathbb{K}[x, y]\) are “compatible” in various ways. For now, let us just note:
Lemma 3.3.1. For \( f \in \mathbb{K}[x, y] \), we have
\[
Z(f \cdot [A, t]) = [A, t]^{-1} \cdot Z(f)
\]
\[
= \{ [A^{-1}, -A^{-1}t](x) : x \in Z(f) \}.
\]

Proof. A point \( x \in \mathbb{K}^2 \) is in \( Z(f \cdot [A, t]) \) iff \( f(Ax + t) = 0 \), which is true iff \( Ax + t = [A, t](x) \) is in \( Z(f) \), which is true iff
\[
x = [A, t]^{-1} \cdot ([A, t]x) \in [A, t]^{-1}Z(f).
\]

In other words, the zero locus of \( f \cdot [A, t] \) is obtained from the zero locus of \( f \) by applying the inverse of the affine transformation \([A, t] \) to \( \mathbb{K}^2 \). The reader should be familiar with the appearance of the “inverse” here, as this also occurs (for the same reason) when one thinks about, say, the relationship between the group of \( f(x) \) and the graph of \( f(ax + b) \), with \( a \in \mathbb{R}^+ \), \( b \in \mathbb{R} \).

Proposition 3.3.2. The degree of a polynomial is invariant under the action of affine transformations. That is,
\[
\deg f = \deg(f \cdot [A, t])
\]
for every \( f \in \mathbb{K}[x, y] \) and every \([A, t] \in \text{Aff}(\mathbb{K}^2)\).

Proof. Consider a degree \( d \) polynomial \( f(x, y) = \sum_{i+j \leq d} a_{ij} x^i y^j \). Then if we write out \( A \) and \( t \) as above, we have
\[
f \cdot [A, t] = \sum_{i+j \leq d} a_{ij} (ax + by + t_1)^i(cx + dy + t_2)^j.
\]
From the rule for Multinomial Expansion and the fact that the degree of a product is the sum of the degrees of the factors, it is clear that (when \( i + j \leq d \)) the polynomial \((ax + by + t_1)^i(cx + dy + t_2)^j\) clearly has degree at most \( d \). Since \( f \cdot [A, t] \) is a sum of such things, we see that the degree of \( f \cdot [A, \bar{s}] \) cannot be any bigger than \( d = \deg f \):
\[
\deg(f \cdot [A, t]) \leq \deg f.
\]
But now we apply this same observation with the polynomial \( f \) replaced by the polynomial \( f \cdot [A, t] \), and \([A, t] \) replaced by its inverse to find the opposite inequality (here we use the fact that we have an action to know that \( f = (f \cdot [A, t]) \cdot [A, t]^{-1} \)).

Definition 3.3.3. Two polynomials \( f, g \in \mathbb{K}[x, y] \) are called affine equivalent iff there is an affine transformation \([A, t] \in \text{Aff}(\mathbb{K}^2) \) and a non-zero scalar \( \lambda \in \mathbb{K}^* \) such that \( \lambda g = f \cdot [A, t] \). When this is the case, one says that \( g \) is obtained from \( f \) by affine linear change of variables (and rescaling). Similarly, we say that two subsets \( S, T \subseteq \mathbb{K}^2 \) are affine equivalent iff \( S = [A, t](T) \) for some affine transformation \([A, t] \in \mathbb{K}^2 \).
One can check, using the fact that our construction of $f \cdot [A, t]$ is an action of $\text{Aff}(\mathbb{K}^2)$ on $\mathbb{K}[x, y]$, that “being affine equivalent” is indeed an equivalence relation (Exercise 3.4). Notice that Lemma 3.3.1 (plus the fact that rescaling a polynomial doesn’t change its zero locus) implies that if $f, g \in \mathbb{K}[x, y]$ are affine equivalent, then their zero loci $Z(f)$, $Z(g)$ are affine equivalent as subsets of $\mathbb{K}^2$. The converse is certainly not true: we will soon see that there are inequivalent degree two polynomials $f, g \in \mathbb{R}[x, y]$ such that both $Z(f)$ and $Z(g)$ are empty!

A very ambitious goal would now be to classify all polynomials and all plane curves up to affine equivalence. (That is, to “explicitly” describe all affine equivalence classes.) To attempt this, one has to think of many “affine invariants,” which is to say, “quantities associated to a polynomial $f \in \mathbb{K}[x, y]$ (or a subset $C \subseteq \mathbb{K}^2$) that depend only on the affine equivalence class of $f$ (or $C$).” Proposition 3.3.2 gives us one such quantity, namely the degree of a polynomial. This lets us break up our classification problem degree by degree.

As a warmup, let us classify linear (that is, degree one) polynomials up to affine equivalence. The corresponding classification of plane curves will be the classification of lines, in the sense of Definition 3.1.4. We will again take up the subject of lines, in the projective setting, in §5.4. We claim that in fact every degree one polynomial is equivalent to the polynomial $x$. A degree one polynomial is a polynomial of the form $f(x, y) = Ax + By + C$ for some real numbers $A, B, C$ with $(A, B) \neq (0, 0)$. After applying the (invertible linear) transformation $(x, y) \mapsto (y, x)$ if necessary, we see that $f$ is equivalent to a polynomial $g$ of the same form where in fact $A \neq 0$. By applying the invertible linear transformation $(x, y) \mapsto (A^{-1}x, y)$, we see that $g$ is equivalent to a polynomial $h$ of the form $x + By + C$. After applying the invertible linear transformation $(x, y) \mapsto (x - By, y)$, we see that $h$ is equivalent to a polynomial $j$ of the form $x + C$. Finally, applying the translation $(x, y) \mapsto (x - C, y)$, we see that $j$ is equivalent to $x$. We leave it as an easy exercise for the reader to see that “a linear polynomial $f(x, y)$ can be recovered (up to multiplication by a non-zero real number) from the corresponding plane curve (line) $Z(f)$”.

We should also emphasize at this point that our notion of affine equivalence is not the only “reasonable” one: There are very legitimate reasons for rejecting the idea that plane curves differing by affine transformation are “equivalent.” For one thing, affine transformations of $\mathbb{R}^2$ do not always distances, or even angles. It would be quite reasonable to try to classify conics in $\mathbb{R}^2$ up to isometry (§1.2). While this can be done in an elementary manner, it is rather tedious and we shall not pursue this here. We will see this issue (the difference between affine equivalence and isometry-equivalence) manifested in our classification of conics in various ways: all ellipses are equivalent to the unit circle, all pairs of intersecting lines are equivalent (regardless of the angle of intersection), and so forth.

### 3.4 Five points determine a conic

If $P$ and $Q$ are distinct points of $\mathbb{K}^2$, then it is easy to see that there is a unique line $L \subseteq \mathbb{K}^2$ for which $P, Q \in L$. Here we prove an analogous statement for conics. The
technique used in the proof will be recycled many times later.

Proposition 3.4.1. Suppose $P_1, \ldots, P_5 \in \mathbb{K}^2$ are any five points, no three of which are collinear. Then, up to multiplying by a non-zero real number, there is a unique degree two polynomial $f(x,y) \in \mathbb{K}[x,y]$ such that $P_1, \ldots, P_5 \in Z(f)$.

Proof. Let us first treat the special case where $P_1 = (0,0)$, $P_2 = (1,0)$, and $P_3 = (0,1)$. The general degree two polynomial takes the form

$$f(x,y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F,$$

where $(A,B,C) \neq (0,0,0)$. The condition that $(0,0) \in Z(f)$ is equivalent to $F = 0$, and the conditions that $(1,0) \in Z(f)$ and $(0,1) \in Z(f)$ are then equivalent to $D = -A$ and $E = -C$, so our $f$ must take the form

$$f(x,y) = Ax^2 + Bxy + Cy^2 - Ax - Cy.$$

If we write $P_4 = (x_1,y_1)$, $P_5 = (x_2,y_2)$, then the conditions that $P_4, P_5 \in Z(f)$ are equivalent to the system of linear equations

$$Ax_1(x_1 - 1) + Bx_1y_1 + Cy_1(y_1 - 1) = 0$$

$$Ax_2(x_2 - 1) + Bx_2y_2 + Cy_2(y_2 - 1) = 0.$$  

By basic linear algebra, we can conclude that this system has a unique non-zero solution $(A,B,C)$, up to scaling, provided that we can show the coefficient vectors

$$(x_1(x_1 - 1), x_1y_1, y_1(y_1 - 1)) \text{ and } (x_2(x_2 - 2), x_2y_2, y_2(y_2 - 1))$$

are linearly independent.

Suppose not. Then there is some $\lambda \in \mathbb{K}$ so that

$$\lambda(x_1(x_1 - 1), x_1y_1, y_1(y_1 - 1)) = (x_2(x_2 - 2), x_2y_2, y_2(y_2 - 1)).$$

Now, since $P_4$ isn’t on the line joining $(0,0)$ and $(1,0)$ (the $x$-axis), we have $y_1 \neq 0$. Since $P_4$ isn’t collinear with $(0,0)$ and $(1,0)$, we have $x_1 \neq 0$. For similar reasons, we have $x_2, y_2 \neq 0$. We can then solve for $\lambda$ using equality of the second coordinates above to find $\lambda = (x_2y_2)/(x_1y_1)$. Plugging in for $\lambda$ and cancelling the non-zero real number $x_2$, equality of the first coordinates above then yields:

$$y_2(x_1 - 1) = y_1(x_2 - 1).$$

But $y(x_1 - 1) = y_1(x - 1)$ is the equation defining the line containing $P_4 = (x_1,y_1)$ and $P_2 = (1,0)$, so this latter equality shows that $P_2, P_4$, and $P_5$ are collinear—a contradiction.

We deduce the general case from the special case as follows: By Proposition 1.4.3 we can find an affine transformation $[A,t] : \mathbb{K}^2 \to \mathbb{K}^2$ taking $P_1, P_2, \text{ and } P_3$ to $(0,0)$, $(1,0)$, and $(0,1)$, respectively. Since an affine transformation takes lines to lines, the five points

$$(0,0) = [A,t](P_1), (1,0) = [A,t](P_2), (0,1) = [A,t](P_3), [A,t](P_4), [A,t](P_5)$$

are five points, no three of which are collinear, and up to scaling, provided that we can show the coefficient vectors
are in general position (no three collinear), so by the special case treated above, there
is a unique (up to rescaling) degree two \( f \in K[x,y] \) such that \( Z(f) \) contains these five
points. By Proposition 3.3.2 and Lemma 3.3.1, the polynomial \( f \cdot [A,t] \) is of degree two
and its zero locus contains \( P_1, \ldots, P_5 \). If \( g \) is any other degree two polynomial whose
zero locus contains \( P_1, \ldots, P_5 \), then that proposition and lemma imply that \( g \cdot [A,t]^{-1} \) is
a degree two polynomial whose zero locus contains the five points listed above, hence
\( g \cdot [A,t]^{-1} \) must be a rescaling of \( f \) by the special case treated above, and hence \( g \)
must be a rescaling of \( f \cdot [A,t] \).

3.5 Plane conics up to affine equivalence

Having dispensed with the classification of degree one polynomials up to affine equivalence
in §3.3, we now turn to the case of degree two polynomials. Unlike the classification of
linear polynomials, this classification is quite sensitive to the field \( K \) over which we work.
The classification over the real numbers \( \mathbb{R} \) is as follows:

**Theorem 3.5.1.** Every polynomial \( f(x,y) \in \mathbb{R}[x,y] \) of degree two is affine equivalent
(Definition 3.3.3) to exactly one of the following:

1. \( x^2 + y^2 - 1 \)
2. \( x^2 - y^2 - 1 \)
3. \( y - x^2 \)
4. \( xy \)
5. \( x^2 \)
6. \( x(x - 1) \)
7. \( x^2 + y^2 \)
8. \( x^2 + y^2 + 1 \)
9. \( x^2 + 1 \)

Over the complex numbers \( \mathbb{C} \), the classification is simpler:

**Theorem 3.5.2.** Every polynomial \( f(x,y) \in \mathbb{C}[x,y] \) of degree two is affine equivalent to
exactly one of the following:

1. \( x^2 + y^2 - 1 \)
2. \( y - x^2 \)
3. \( xy \)
4. \( x^2 \)
5. \( x(x - 1) \)
We are not going to prove these theorems, though we will prove their “projective analogues” in §5.6. The proofs are elementary: the basic technique, which will also be used in the proofs our projective analogues of these theorems, amounts to little more than the idea of “completing the square” familiar from calculus (or from the derivation of the Quadratic Formula, say). However, the proofs in the affine case are slightly more tedious, as there is more case-by-case analysis. The projective analogues of this theorem are also in many ways more important and fundamental, as they are related to the problem of classifying symmetric bilinear forms, which is important in linear algebra.

From the classification of degree two polynomials, we can easily obtain the classification of the conics given by their zero loci:

**Corollary 3.5.3.** For every polynomial $f(x,y) \in \mathbb{R}[x,y]$ of degree two, the corresponding plane curve $Z(f) \subseteq \mathbb{R}^2$ is either empty, or affine equivalent to exactly one of the following:

1. the circle $Z(x^2 + y^2 - 1)$
2. the hyperbola $Z(x^2 - y^2 - 1)$
3. the parabola $Z(y - x^2)$
4. the pair of intersecting lines $Z(xy)$
5. the “doubled” line $Z(x^2)$
6. the pair of parallel lines $Z(x(x - 1))$
7. the single point $Z(x^2 + y^2)$

**Proof.** This will follows from Theorem 3.5.1 and Lemma 3.3.1 once we prove that no two of the non-empty plane curves on the list are affine equivalent. (A priori, it is quite conceivable that two inequivalent polynomials determine equivalent plane curves, and, indeed, this is the case—the two inequivalent polynomials $x^2 + y^2 + 1$ and $x^2 + 1$ both determine the empty curve, but, what we are about to argue is that, in fact, this does not happen, except in the aforementioned “empty curves” case.)

This can be done in various ways. Perhaps the most direct way is to rule out the equivalence of different curves on the list on elementary topological grounds: affine transformations are, in particular, homeomorphisms, and the point is that very few pairs of curves on the list are even homeomorphic. (The reader who doesn’t know what “homeomorphism” means should just take this as a precise way of making precise the obvious “differences” we observe such as the fact that the parabola and circle are “connected,” whereas the hyperbola “has two components.” Similarly, the circle is “bounded” (compact), while the parabola and hyperbola are not.) The exceptions are: the hyperbola is homeomorphic to the two parallel lines (both are just disjoint unions of two copies of the real line), and the parabola is homeomorphic to the “double” line (both are homeomorphic to the real line). But these pairs of plane curves are easily seen to be distinct (inequivalent) on the grounds that affine transformations take lines to lines. For example, one need only write down three non-colinear points on the parabola to know that it cannot be equivalent to the “double” line. 

\[\square\]
3.6 Invariants of affine conics

The most difficult thing to do in a classification problem is often to prove that various things are not equivalent. To do this, one typically tries to attach some kind of invariant—a number, say—to the objects being classified in such a way that this invariant depends only on the equivalence class of the object.

The relevant invariants for plane conics are the discriminant and the degeneracy. Consider a typical polynomial

\[ f(x, y) = Ax^2 + Bxy + Cy^2 + Dx + Ey + F \]  

(3.2)
of degree at most two with coefficients in a field \( K \) where \( 2 \in K^* \) (for example: \( K = \mathbb{Q}, \mathbb{R}, \) or \( \mathbb{C} \)). A key point is to note that \( f(x, y) \) can be described in terms of matrix multiplication, as follows. We let \( \mathcal{M}(f) \) be the symmetric \( 3 \times 3 \) matrix with entries in \( K \) defined by

\[ \mathcal{M}(f) := \begin{pmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{pmatrix}. \]  

(3.3)

(Here we need to assume \( 2 \in K^* \).) Then \( f(x, y) \) and the matrix \( \mathcal{M}(f) \) are related via the formula:

\[ f(x, y) = (x, y, 1) \begin{pmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \]  

(3.4)

We call the matrix \( \mathcal{M}(f) \) the matrix associated to \( f \). Given any symmetric \( 3 \times 3 \) matrix \( M \) with entries in \( K \), we can define a polynomial \( f(M) \in K[x, y] \) of degree at most two by setting

\[ f(M) = (x, y, 1)^t \mathcal{M}(x, y, 1). \]

It is clear from formula (3.4) that \( f \mapsto \mathcal{M}(f) \) establishes a bijection from polynomials in \( x, y \) of degree at most two to symmetric \( 3 \times 3 \) matrices with inverse \( M \mapsto f(M) \).

**Definition 3.6.1.** For \( f \) as above, we define the discriminant \( \mathcal{D}(f) \in K \) and the degeneracy \( \Delta(f) \in K \) of \( f \) by

\[ \mathcal{D}(f) := B^2 - 4AC \]
\[ \Delta(f) := \det \mathcal{M}(f). \]

We say that \( f(x, y) \) is degenerate if \( \Delta(f) = 0 \) and non-degenerate otherwise.

The behaviour of \( \mathcal{D}(f) \) and \( \Delta(f) \) under affine transformations is as follows.
Lemma 3.6.2. For a polynomial \( f(x, y) \in \mathbb{K}[x, y] \) of degree at most two, an affine transformation \([A, \vec{t}] \in \text{Aff}(\mathbb{K}^2)\), and \( \lambda \in \mathbb{K} \), we have

\[
D(\lambda f \cdot [A, \vec{t}]) = \lambda^2 (\det A)^2 D(f) \\
\Delta(\lambda f \cdot [A, \vec{t}]) = \lambda^3 (\det A)^2 \Delta(f).
\]

Proof. The most direct proof is to just compute everything explicitly. The behaviour of \( D \) and \( \Delta \) under multiplying by \( \lambda \) is easy to understand (multiplying every entry of a \( 3 \times 3 \) matrix by \( \lambda \) changes its determinant by \( \lambda^3 \)), so it is enough to deal with the case \( \lambda = 1 \). Suppose

\[
A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \vec{t} = \begin{pmatrix} s \\ t \end{pmatrix}
\]

and \( f \) is given by (3.2). Then

\[
f \cdot [A, \vec{t}] = f(ax + by + s, cx + dy + t) \\
= A'x^2 + B'xy + C'y^2 + D'x + E'y + F',
\]

where

\[
A' = Aa^2 + Bac +Cc^2 \\
B' = 2Aab + B(ad + bc) + 2Ccd \\
C' = Ab^2 + Bbd + Cd^2 \\
D' = 2Aas + B(at + sc) + 2Cct + Da + Ec \\
E' = 2Abs + B(bt + ds) + 2Cdt + Db + Ed \\
F' = As^2 + Bst + Ct^2 + Ds + Et + F.
\]

Now a certain amount of laborious (but elementary) calculation yields

\[
D(f \cdot [A, \vec{t}]) = (B')^2 - 4A'C' \\
= (ad - bc)^2 (B^2 - 4AC) \\
= \det(A)^2 D(f) \\
\Delta(f \cdot [A, \vec{t}]) = \det \begin{pmatrix} A' & B'/2 & D'/2 \\ B'/2 & C' & E'/2 \\ D'/2 & E'/2 & F' \end{pmatrix} \\
= (ad - bc)^2 \det \begin{pmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{pmatrix} \\
= (\det A)^2 \Delta(f).
\]
A slightly more sophisticated treatment of $\Delta(f)$ using the matrix $M(f)$ proceeds as follows. We calculate

$$f \cdot [A, \overline{t}] = (ax + by + s, cx + dy + t) M(f) \begin{pmatrix} ax + by + s \\ cx + dy + t \\ 1 \end{pmatrix}$$

$$= (x, y, 1) \begin{pmatrix} a & c & 0 \\ b & d & 0 \\ s & t & 1 \end{pmatrix} M(f) \begin{pmatrix} a & b & s \\ c & d & t \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

so we see that the matrix

$$M(f \cdot [A, \overline{t}]) = \begin{pmatrix} A' & B'/2 & D'/2 \\ B'/2 & C' & E'/2 \\ D'/2 & E'/2 & F' \end{pmatrix}$$

corresponding to $f \cdot [A, \overline{s}]$ is given by

$$M(f \cdot [A, \overline{t}]) = \begin{pmatrix} a & b & s \\ c & d & t \\ 0 & 0 & 1 \end{pmatrix}^{tr} M(f) \begin{pmatrix} a & b & s \\ c & d & t \\ 0 & 0 & 1 \end{pmatrix}.$$ 

The second formula of the lemma now follows from standard properties of determinants (invariance under transpose, determinant of a product is the product of the determinants) using the trivial calculation

$$\det \begin{pmatrix} a & b & s \\ c & d & t \\ 0 & 0 & 1 \end{pmatrix} = \det A.$$

\[\square\]

**Corollary 3.6.3.** Let $\mathbb{K}$ be a field with $2 \in \mathbb{K}^\ast$ and let $f \in \mathbb{K}[x, y]$ be a polynomial of degree 2. Then $\Delta(f) = 0$ iff $\Delta(g) = 0$ for some (equivalently any) $g$ affine equivalent to $f$. Similarly, when $\mathbb{K} = \mathbb{R}$, the “sign” of $Df$ in $\{+, -, 0\}$ depends only on the affine equivalence class of $f$.

We can use $D$ and $\Delta$ to distinguish between most of the pairs of plane curves on the list in Theorem 3.5.1. The values of $D$ and $\Delta$ for the curves on that list are tabulated.
below:

\[
\begin{array}{ccc}
\quad f(x, y) & D(f) & \Delta(f) \\
x^2 + y^2 - 1 & (-) & \neq 0 \\
x^2 - y^2 - 1 & (+) & \neq 0 \\
y - x^2 & (0) & \neq 0 \\
xy & (+) & 0 \\
x^2 & (0) & 0 \\
x(x - 1) & (0) & 0 \\
x^2 + y^2 & (-) & 0 \\
x^2 + y^2 + 1 & (\ -\ ) & \neq 0 \\
x^2 + 1 & (0) & 0 \\
\end{array}
\]

At this point, we shouldn’t be shy about admitting that the definitions of the discriminant and the \(\Delta\)-invariant seem to “come out of the blue.” Later, in Proposition 5.5.3, we will give a completely “geometric” meaning of the discriminant.

### 3.7 Exercises

**Exercise 3.1.** Suppose \(\mathbb{K}\) is an algebraically closed field (take \(\mathbb{K} = \mathbb{C}\) if you want) and \(f \in \mathbb{K}[x, y]\) is a non-constant polynomial. Prove that \(Z(f) \neq \emptyset\).

**Exercise 3.2.** Suppose \(f \in \mathbb{K}[x, y]\) is a non-constant polynomial and \(n\) is an integer greater than one. Prove that every point of \(Z(f^n)\) is a singular point. More generally, prove that if \(f = gh\) with \(g\) and \(h\) non-constant, then any point of \(Z(g) \cap Z(h)\) is a singular point of \(Z(f)\).

**Exercise 3.3.** Let \(f \in \mathbb{K}[x, y]\) be a polynomial, \([A, t] \in \text{Aff}(\mathbb{K}^2)\) an affine transformation. Prove that a point \(P \in Z(f)\) is a singular point of \(Z(f)\) iff \([A, t]^{-1}(P)\) is a singular point of \(Z(f \cdot [A, t])\).

**Exercise 3.4.** Prove that being affine equivalent (Definition 3.3.3) is an equivalence relation on \(\mathbb{K}[x, y]\).

**Exercise 3.5.** Consider the degree two polynomials \(f := x^2 + y^2, g := -x^2 - y^2\) in \(\mathbb{R}[x, y]\). Show that there is no \(A \in \text{Aff}(\mathbb{R}^2)\) such that \(g = f \cdot A\). (Nevertheless, we do consider \(f\) and \(g\) to be affine equivalent because \((-1)f = g\).)

**Exercise 3.6.** For each of the following polynomials \(f \in \mathbb{R}[x, y]\) decide which of the five polynomials in the list of Theorem 3.5.1 is affine equivalent to \(f\):

1. \(zy - x^2\)
2. \(xy\)
3. \(x^2 + xy + z^2\)
4. \(2x^2 + 2xz + z^2 + 2xy + y^2\)
Exercise 3.7. For $f \in \mathbb{R}[x, y]$, let $\text{Stab } f$ be the set of $A \in \text{Aff}(\mathbb{R}^2)$ such that $f \cdot A = f$.

1. Show that $\text{Stab } f$ is a subgroup of $\text{Aff}(\mathbb{R}^2)$—i.e. show that $\text{Id} \in \text{Stab } f$, that $A^{-1} \in \text{Stab } f$ whenever $A \in \text{Stab } f$, and that $AA' \in \text{Stab } f$ whenever $A, A' \in \text{Stab } f$.

2. Note that $\text{Stab } f = \text{Stab } (\lambda f)$ for any non-zero $\lambda \in \mathbb{R}$. Show that if $f$ and $g$ are Aff-equivalent, then $\text{Stab } f$ is conjugate to $\text{Stab } g$ in $\text{Aff}(\mathbb{R}^2)$—i.e. there is some $A \in \text{Aff}(\mathbb{R}^2)$ such that

$$\text{Stab } g = \{ A^{-1}A' : A' \in \text{Stab } f \}.$$ 

3. Show that $\text{Stab}(x^2 + y^2) = \text{Stab}(x^2 + y^2 - 1)$ is equal to the orthogonal group $O_2 < \text{Aff}(\mathbb{R}^2)$.

4. Calculate $\text{Stab } f$ for $f = y - x^2$. Do you recognize this group? (Forget about the description of it as a subgroup of $\text{Aff}(\mathbb{R}^2)$.)
Chapter 4

Projective Spaces

In this brief chapter we shall introduce and study the projective spaces $\mathbb{K}P^n$ for a field $\mathbb{K}$. The idea of the construction of $\mathbb{K}P^n$ is to enlarge $\mathbb{K}^n$ by adding some additional "points at infinity," one for each line through the origin in $\mathbb{K}^n$. One surprising feature of the construction is that the newly added points "at infinity" are actually "on the same footing" as the old "finite" points. We shall make this precise in §4.4 by introducing and studying the projective general linear groups $\text{PGL}_n(\mathbb{K})$. The group $\text{PGL}_n(\mathbb{K})$ acts on $\mathbb{K}P^n$ through projective transformations in a manner similar to the way $\text{Aff}(\mathbb{K}^n)$ acts on $\mathbb{K}^n$. Indeed, we shall see that $\text{Aff}(\mathbb{K}^n)$ can be identified with the subgroup of $\text{PGL}_n(\mathbb{K})$ consisting of those $\mathcal{A} \in \text{PGL}_n(\mathbb{K})$ such that the associated map $\mathcal{A}: \mathbb{K}P^n \to \mathbb{K}P^n$ takes the locus of "finite points" $\mathbb{K}^n \subseteq \mathbb{K}P^n$ into itself.

In §4.2 we discuss linear subspaces of projective spaces—this concept leads to the definition of a line and to notions of "general position" which we be important later.

Special attention will be given to the real projective spaces $\mathbb{R}P^n$ and the complex projective spaces $\mathbb{C}P^n$ in §4.3. A full appreciation of this section requires some background in topology—this material is not strictly necessary elsewhere in the text.

4.1 Motivation and construction

Before we give the general definition of the projective spaces $\mathbb{K}P^n$, let us motivate the construction by discussing $\mathbb{R}P^1$.

One “problem” with the real line $\mathbb{R}$ is that it is not compact. The general notion of compactness is a concept of topology with which we shall not particularly concern ourselves here, but in this case, let us rephrase the issue a bit by noting that there are sequences of points $a_1, a_2, \ldots$ in $\mathbb{R}$ which do not have a convergent subsequence. For example, $1, 2, 3, \ldots$ is such a sequence. Now, the astute reader with some calculus background would probably suggest that the limit of this sequence should be “$\infty$,” so let us try to make some sense of this. We are led to attempt to enlarge the real line $\mathbb{R}$ by adding one point $\infty$, thus obtaining a (slightly) larger space $\mathbb{R}P^1 = \mathbb{R} \bigcup \{\infty\}$. But now we have to decide how to “topologize” this larger set $\mathbb{R}P^1$. In other words, we have to give some precise definition of when a sequence of points $a_1, a_2, \ldots$ in $\mathbb{R}P^1$ “converges to $\infty$.” A very reasonable way to do this is to agree that $a_1, a_2, \ldots$ converges to $\infty$ iff the
sequence $1/a_1, 1/a_2, \ldots$ converges to 0 (in the usual epsilon-delta sense), where we agree that $1/\infty := 0$.

With this understanding, let us prove that any sequence $a_1, a_2, \ldots$ in $\mathbb{R}P^1$ has a convergent subsequence. First of all, if infinitely many of the $a_i$ are equal to $\infty$, then those $a_i$ of course give a subsequence converging to $\infty$. On the other hand, if only finitely many of the $a_i$ are equal to $\infty$, then we discard them to obtain a subsequence contained in $\mathbb{R} \subseteq \mathbb{R}P^1$. We thus reduce to the case where our sequence $a_1, a_2, \ldots$ is contained in $\mathbb{R} \subseteq \mathbb{R}P^1$. Now, if the magnitudes $|a_i|$ of the $a_i$ are bounded, say by $M$, then we can find a convergent subsequence by basic results from calculus or topology (because the closed interval $[-M, M]$ is compact). On the other hand, if the magnitudes $|a_i|$ are unbounded, then we can find an increasing sequence of integers

$$0 < i_1 < i_2 < \cdots$$

so that $|a_{i_j}| > j$ for each $j \in \{1, 2, \ldots\}$. The subsequence $a_{i_1}, a_{i_2}, \ldots$ of $a_1, a_2, \ldots$ then converges to $\infty$ because $|1/a_{i_j}| < 1/j$ for each $j \in \{1, 2, \ldots\}$.

There are many other ways to think about this construction of $\mathbb{R}P^1$. For example, we can view $\mathbb{R}P^1$ as being obtained from two copies $\mathbb{R}_1, \mathbb{R}_2$ of the real line $\mathbb{R}$ by identifying a non-zero point $x \in \mathbb{R}_1$ with the (non-zero) point $1/x \in \mathbb{R}_2$. Every point in the resulting “glued up” space will have a unique representative in $\mathbb{R}_1$, with the exception of the origin $0 \in \mathbb{R}_2$, which plays the role of $\infty$ in the previous construction.

Another way to think about the construction of $\mathbb{R}P^1$ more in line with what we will do in general, is to notice that any point $x \in \mathbb{R}$ determines a line $L_x$ passing through the origin in $\mathbb{R}^2$—namely, the line with slope $x$. Notice that every line through the origin in $\mathbb{R}^2$ will be equal to $L_x$ for a unique real number $x$ with one exception: the vertical line $L_\infty$ through the origin, which we might reasonably describe as the “line through the origin with slope $\infty$.” The reader with some background in linear algebra will no doubt realize that a line through the origin in $\mathbb{R}^2$ is the same thing as a one dimensional subspace of the vector space $\mathbb{R}^2$. This is the point of view that will generalize well.

Yet another way to think about $\mathbb{R}P^1$, also in line with what we will do later, is to think of a real number $x$ as a ratio of two real numbers $r = x/1 = 2x/2 = \cdots$ where the denominator is non-zero. With this understanding, we might reasonable agree that $\infty$ can also be viewed as a ratio of two real numbers: $\infty = 1/0 = 5/0 = \cdots$, where the numerator is non-zero. To make this a bit more formal, we can consider the set $\mathbb{R}^2 \setminus \{(0, 0)\}$ of pairs of real numbers $(x, y)$, not both zero, and impose the equivalence relation $\sim$ where we declare $(x_1, y_1) \sim (x_2, y_2)$ iff $x_1y_2 = y_1x_2$. Notice that, if $y_1$ and $y_2$ are non-zero, we have $(x_1, y_1) \sim (x_2, y_2)$ iff the real numbers $x_1/y_1$ and $x_2/y_2$ are equal. Denote the equivalence class of $(x, y)$ by $[x : y]$. Notice that every such equivalence class is of the form $[x : 1]$ for a unique real number $x$ with one exception: the equivalence class $[1 : 0]$ (which is equal to the equivalence class $[x : 0]$ for any non-zero real number $x$).

With all this motivation, we are led to the general definition of the projective spaces $\mathbb{K}P^n$:

**Definition 4.1.1.** For a field $\mathbb{K}$, we let $\mathbb{K}P^n$ denote the set of one dimensional subspaces of the $\mathbb{K}$-vector space $\mathbb{K}^{n+1}$. We call $\mathbb{K}P^n$ the n-dimensional projective space over $\mathbb{K}$. For a non-zero vector $x \in \mathbb{K}^{n+1} \setminus \{0\}$, we let $[x] \in \mathbb{K}P^n$ denote the span of $x$. If
With this understanding, we can also view $x$ the inverse of infinite points $y$ with Definition 4.1.3. Define a bijection the single point of $\mathbb{K}P$ $K$ in (one dimensional linear subspaces of) $\mathbb{K}P$ $K$ by identifying $\psi(U)$ $x$ $K$ and $y$ $K$ have identified the set of points at infinity in $\mathbb{K}P$. We call $\mathbb{P}(\mathbb{V})$ the projectivization of $\mathbb{V}$.

Notice that every point of $\mathbb{K}P^n$ is of the form $[x]$ for some $x \in \mathbb{K}^{n+1} \setminus \{0\}$ (because any one dimensional subspace of a vector space is spanned by any non-zero element of it) and we have $[x] = [y]$ iff $x = \lambda y$ for some $\lambda \in \mathbb{K}^*$. We can thus view $\mathbb{K}P^n$ as the set of equivalence classes in $\mathbb{K}^{n+1} \setminus \{0\}$ where we declare $x \sim y$ iff $x = \lambda y$ for some $\lambda \in \mathbb{K}^*$. With this understanding, we can also view $[x]$ as denoting the $\sim$-equivalence class of $x \in \mathbb{K}^{n+1} \setminus \{0\}$.

If $(x_0, \ldots, x_n) = \lambda(y_0, \ldots, y_n)$ for $\lambda \in \mathbb{K}^*$, then clearly $x_i = 0$ iff $y_i = 0$, so the question of whether $x_i = 0$ for a point $x = (x_0, \ldots, x_n) \in \mathbb{K}^{n+1} \setminus \{0\}$ depends only on the equivalence class $[x]$ of $x$. Hence we can make the following:

**Definition 4.1.2.** Let $U_0, \ldots, U_n \subseteq \mathbb{K}P^n$ be the subsets defined by

$$U_i := \{[x_0 : \cdots : x_n] \in \mathbb{K}P^n : x_i \neq 0\}.$$ 

The subset $U_n$ of $\mathbb{K}P^n$ will be called the set of finite points of $\mathbb{K}P^n$, and its complement $\mathbb{K}P^n \setminus U_n = \{[x_0 : \cdots : x_n] \in \mathbb{K}P^n : x_i = 0\}$ will be called the set of infinite points of $\mathbb{K}P^n$.

Note that $\mathbb{K}P^n = U_0 \cup \cdots \cup U_n$.

**Definition 4.1.3.** Define a bijection $\phi_i : \mathbb{K}^n \to U_i \subseteq \mathbb{K}P^n$ by

$$\phi_i : \mathbb{K}^n \to \mathbb{K}P^n$$

$$(a_1, \ldots, a_n) \mapsto [a_1 : \cdots : a_{i-1} : 1 : a_i : \cdots : a_n].$$

The inverse $\phi_i^{-1}$ of $\phi_i$ is given by

$$\phi_i^{-1}[x_0 : \cdots : x_n] = (x_0/x_i, \ldots, x_{i-1}/x_i, x_{i+1}/x_i, \ldots, x_n/x_i).$$

Similarly, we let $\psi_i : \mathbb{K}P^{n-1} \to \mathbb{K}P^n \setminus U_i$ be the bijection defined by

$$\psi_i : \mathbb{K}P^{n-1} \to \mathbb{K}P^n \setminus U_i$$

$$[x_0 : \cdots : x_{n-1}] \mapsto [x_0 : \cdots : x_{i-1} : 0 : x_i : \cdots : x_{n-1}].$$

In particular, $\phi_n$ gives a bijection between $\mathbb{K}^n$ and the set $U_n$ of finite points of $\mathbb{K}P^n$. We often suppress notation for this bijection and simply regard $\mathbb{K}^n$ as a subset of $\mathbb{K}P^n$ by identifying $\mathbb{K}^n$ with $U_n$ via $\phi_n$. In other words, we don’t make much distinction between a point $(x_1, \ldots, x_n) \in \mathbb{K}^n$ and the “corresponding” point $[x_1 : \cdots : x_n : 1]$ of $U_n \subseteq \mathbb{K}P^n$. Similarly, we identify $\mathbb{K}P^{n-1}$ with the set of infinite points in $\mathbb{K}P^n$ via the bijection $\psi_i$, so if there is no chance of confusion we do not make any distinction between $[x_0 : \cdots : x_{n-1}] \in \mathbb{K}P^{n-1}$ and $[x_0 : \cdots : x_{n-1} : 0] \in \mathbb{K}P^n$. In particular, notice that we have identified the set of points at infinity in $\mathbb{K}P^n$ with the set of lines through the origin in (one dimensional linear subspaces of) $\mathbb{K}^n = U_n \subseteq \mathbb{K}P^n$.

For example, in the case $n = 1$, the “finite points” of $\mathbb{K}P^1$ are those of the form $[x : y]$ with $y \neq 0$, and there is a single infinite point $[1 : 0]$, often denoted $\infty$, corresponding to the single point of $\mathbb{K}P^0$. 

\[ x = (x_0, \ldots, x_n) \in \mathbb{K}^{n+1}, \text{ we write } [x_0 : \cdots : x_n] \text{ as an alternative to } [x]. \]
4.2 Linear subspaces of projective space

The definition of $\mathbb{P}(V)$ for an abstract $\mathbb{K}$-vector space $V$ is convenient, even if one is only interested in $\mathbb{K}^n$, for essentially the same reasons that it is convenient to have an abstract notion of “vector space” even if one is only interested in the vector space $\mathbb{K}^n$—for example, because “abstract” vector spaces are going to arise naturally as subspaces of $\mathbb{K}^n$. By definition, we have $\mathbb{K}^n = \mathbb{P}(\mathbb{K}^{n+1})$. (The notation is unfortunate.) Notice also that if $V$ is a subspace of a vector space $W$, then $\mathbb{P}(V)$ is a subset of $\mathbb{P}(W)$ because any one dimensional subspace of $V$ is also a one dimensional subspace of $W$. In particular, a subspace $V$ of the vector space $\mathbb{K}^{n+1}$ (“linear subspace of $\mathbb{K}^{n+1}$”) gives rise to a subset $\mathbb{P}(V) \subseteq \mathbb{K}^n$.

**Definition 4.2.1.** A linear subspace of $\mathbb{K}^n$ of dimension $d$ is a subset of $\mathbb{K}^n$ of the form $\mathbb{P}(V)$, for $V$ a linear subspace of $\mathbb{K}^{n+1}$ of dimension $d+1$ (as a $\mathbb{K}$ vector space). A linear subspace of dimension one (resp. two) is often called a line (resp. plane), while a linear subspace of $\mathbb{K}^n$ of dimension $n-1$ (“codimension one”) is often called a hyperplane.

For example, the set of “infinite” points of $\mathbb{K}^n$ is a hyperplane—it is $\mathbb{P}(V)$ where $V$ is the subspace of $\mathbb{K}^{n+1}$ spanned by the first $n$ basis vectors $e_0, \ldots, e_{n-1}$.

**Definition 4.2.2.** Points $P_1, \ldots, P_m \in \mathbb{K}^n$ are said to be in general position iff, for all $k \in \{1, \ldots, \max m, n+1\}$, no $k$ of the $P_i$ are contained in a linear subspace of $\mathbb{K}^n$ of dimension less than $k-1$.

For example, when $n \geq 2$, three points $P_1, P_2, P_3 \in \mathbb{K}^n$ are in general position iff they are not contained in any line. Three points $P_1, P_2, P_3 \in \mathbb{K}^1$ are in general position iff they are distinct.

4.3 Real and complex projective spaces

In order to get a feeling for the general projective spaces $\mathbb{K}^n$ introduced in §4.1 it helps to think about the cases where $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{C}$.

By definition, $\mathbb{R}^n$ is the set of line through the origin in $\mathbb{R}^{n+1}$. Each such line $L$ will intersect the $n$-dimensional sphere

$$S^n := \{ x \in \mathbb{R}^{n+1} : |x| = 1 \}$$

in precisely two points, which are “antipodal” in the sense that one is obtained from the other by multiplying the coordinates by $-1$. Indeed, if we choose any non-zero point $x \in L$, then the two points in question are $x/|x|$ and $-x/|x|$. “Conversely,” if we have a point $x \in S^n \subseteq \mathbb{R}^{n+1}$, then we get a line through the origin $L_x$ by considering the span of $x$. The line $L_x$ intersects $S^n$ at the points $x$ and the antipodal point $-x$. We thus see that $\mathbb{R}^{n+1}$ can be obtained from the $n$-dimensional sphere $S^n$ by identifying antipodal points.

Alternatively, we can consider the upper hemisphere

$$S^n_+ := \{ x = (x_0, \ldots, x_n) \in S^n : x_n > 0 \}$$
of $S^n$. Notice that projecting onto the first $n$ coordinates defines a (continuous) bijection from $S^n$ to the $n$-dimensional disc

$$D^n := \{ x = (x_0, \ldots, x_{n-1}) \in \mathbb{R}^n : |x| \leq 1 \}.$$ 

The inverse of this bijection is given by

$$x = (x_0, \ldots, x_{n-1}) \mapsto (x_0, \ldots, x_{n-1}, \sqrt{1 - |x|^2}).$$

By reasoning as above, we see that each line through the origin in $\mathbb{R}^{n+1}$ will intersect the upper hemisphere $S^n$ in either one point or two points; this intersection will consist of two antipodal points iff the line is contained in the hyperplane $\mathbb{R}^n \subseteq \mathbb{R}^{n+1}$ where the last coordinate $x_n$ is zero. From this perspective we see that $\mathbb{RP}^n$ can be obtained from the $n$-dimensional disc $D^n$ by identifying antipodal points of its boundary sphere $S^{n-1}$.

For example, $\mathbb{RP}^1$ is the quotient of $S^1$ by the antipodal map $x \mapsto -x$, but it can also be described by identifying the two boundary points of the upper half circle (which is homeomorphic to a closed interval), so it is also a circle. The quotient map $S^1 \to \mathbb{RP}^1 = S^1$ can be viewed as the map $z \mapsto z^2$ if we view the circle $S^1$ as the set of complex numbers $z$ of magnitude one.

The set $\mathbb{RP}^2$ can similarly be viewed as the quotient of the usual sphere $S^2 \subseteq \mathbb{R}^3$ obtained by identifying antipodal points, or as the quotient of the unit disc $D^2 \subseteq \mathbb{R}^2$ obtained by identifying antipodal points on the boundary circle $S^1 \subseteq D^2$.

The story for complex projective space $\mathbb{CP}^n$ is similar. For any point $[z] \in \mathbb{CP}^n$, we can find a representative $z \in [z]$ lying in the sphere

$$S^{2n+1} := \{ z \in \mathbb{C}^{n+1} : |z| = 1 \}.$$ 

Two points $z, z' \in S^{2n+1}$ determine the same point $[z] = [z']$ of $\mathbb{CP}^n$ iff $z = \lambda z'$ for some $\lambda \in \mathbb{C}$—in fact, taking $| \cdot |$, we see that $|\lambda| = 1$, so $\lambda$ will lie in the unit circle

$$S^1 = \{ \lambda \in \mathbb{C} : |\lambda| = 1 \}.$$ 

This shows that $\mathbb{CP}^n$ can be described as the quotient of $S^{2n+1}$ by the free action of $S^1$ given by $(\lambda, z) \mapsto \lambda z$.

For the student with some background in topology, we remark that all of the descriptions of $\mathbb{RP}^n$ and $\mathbb{CP}^n$ given in these notes can actually be viewed as descriptions of a topological space (not just a set). In fact all of these descriptions are describing the same topology:

**Definition 4.3.1.** We view $\mathbb{RP}^n$ as a topological space by declaring a subset $U \subseteq \mathbb{RP}^n$ to be open iff the following equivalent conditions are satisfied:

1. The preimage of $U$ under the map

   $$\mathbb{R}^{n+1} \setminus \{0\} \to \mathbb{RP}^n$$

   given by $x \mapsto [x]$ is open in $\mathbb{R}^{n+1}$.

2. The preimage of $U$ under the map $S^n \to \mathbb{RP}^n$ given by $x \mapsto [x]$ is open in $S^n$. 
3. The preimage of $U$ under the map $S^n_+ \to \mathbb{RP}^n$ given by $x \mapsto [x] = \operatorname{Span} x$ is open in $S^n_+$.

4. The preimage of $U \cap U_i$ under the bijection $\phi_i : \mathbb{R}^n \to U_i \subseteq \mathbb{RP}^n$ (Definition 4.1.3) is open for each $i \in \{0, \ldots, n\}$.

(The equivalence of these conditions is Exercise 4.2.)

**Definition 4.3.2.** We topologize $\mathbb{CP}^n$ by declaring a subset $U \subseteq \mathbb{CP}^n$ to be *open* iff the following equivalent conditions are satisfied:

1. The preimage of $U$ under the map $\mathbb{C}^{n+1} \setminus \{0\} \to \mathbb{CP}^n$ given by $z \mapsto [z]$ is open in $\mathbb{C}^{n+1}$.

2. The preimage of $U$ under the map $S^{2n+1} \to \mathbb{CP}^n$ given by $z \mapsto [z]$ is open in $S^{2n+1}$.

3. The preimage of $U \cap U_i$ under the bijection $\phi_i : \mathbb{C}^n \to U_i \subseteq \mathbb{CP}^n$ is open for each $i \in \{0, \ldots, n\}$.

In particular, $\mathbb{CP}^n$ and $\mathbb{RP}^n$ are both quotients of spheres and spheres are compact, so we have:

**Proposition 4.3.3.** For each $n$, real projective space $\mathbb{RP}^n$ and compact projective space $\mathbb{CP}^n$ are compact topological spaces.

In Exercise 4.4 you will show that $\mathbb{CP}^1$ is homeomorphic to the 2-sphere $S^2$, so that the map $S^3 \to \mathbb{CP}^1$ given by $z \mapsto [z]$ discussed above (the quotient of the circle action on $S^3$) can be viewed as a map $f : S^3 \to S^2$ for which $f^{-1}(x)$ is a circle for each $x \in S^2$.

The map $f$ is called the *Hopf fibration*—it plays a role in various branches of topology. It is the most basic example of an “interesting” map from a sphere to a sphere of lower dimension.

### 4.4 Projective transformations

Now we will define the projective analogue of the affine transformations studied in §1.4. First, notice that an invertible linear transformation $A \in \text{GL}_{n+1}(\mathbb{K})$ gives rise to a bijection

$$
\overline{A} : \mathbb{KP}^n \to \mathbb{KP}^n
$$

$$
[x] \mapsto [Ax].
$$

This is well-defined because if $x = \lambda y$ for $\lambda \in \mathbb{K}^*$, then $\lambda Ax = A\lambda x = Ay$, so $[Ay] = [Ax]$. We also need to know that $A$ is invertible to ensure that $Ax \neq 0$ when $x \neq 0$, so that $[Ax]$ makes sense as a point of $\mathbb{KP}^n$. The bijections (4.1) are called *projective transformations*. Clearly the composition $\overline{A} \circ \overline{B}$ of the projective transformations associated to invertible matrices $A, B \in \text{GL}_{n+1}(\mathbb{K})$ is the projective transformation $\overline{AB}$ associated to their
product. In particular, the inverse of the projective transformation \( \overline{A} \) associated to \( A \in GL_{n+1}(\mathbb{K}) \) is the projective transformation \( \overline{A}^{-1} \) associated to its inverse. Notice that for \( \lambda \in \mathbb{K}^* \) and \( A \in GL_{n+1}(\mathbb{K}) \), we have \( \overline{A} = \lambda \overline{A} \) (for the same reason that \( \overline{A} \) is well-defined). We are thus led to consider the following group:

**Definition 4.4.1.** The *projective general linear group* \( \operatorname{PGL}_n(\mathbb{K}) \) is the quotient of \( GL_{n+1}(\mathbb{K}) \) by the (normal—in fact central) subgroup \( \{ \lambda I : \lambda \in \mathbb{K}^* \} \). We denote the image of \( A \in GL_{n+1}(\mathbb{K}) \) in \( \operatorname{PGL}_n(\mathbb{K}) \) by \( \overline{A} \).

Explicitly, two invertible matrices \( A, B \in GL_{n+1}(\mathbb{K}) \) have the same image in \( \operatorname{PGL}_n(\mathbb{K}) \) iff \( A = \lambda B \) for some \( \lambda \in \mathbb{K}^* \).

**Remark 4.4.2.** The indexing of \( \operatorname{PGL}_n(\mathbb{K}) \) varies. Some people might write \( \operatorname{PGL}_{n+1}(\mathbb{K}) \) for what I have called \( \operatorname{PGL}_n(\mathbb{K}) \).

The next result justifies our use of the notation \( \overline{A} \) for both the image of a matrix \( A \in GL_{n+1}(\mathbb{K}) \) in \( \operatorname{PGL}_n(\mathbb{K}) \) and for the associated projective transformation \( \overline{A} : \mathbb{KP}^n \to \mathbb{KP}^n \).

**Proposition 4.4.3.** Two invertible matrices \( A, B \in GL_{n+1}(\mathbb{K}) \) have the same image in \( \operatorname{PGL}_n(\mathbb{K}) \) iff the projective transformations \( \overline{A}, \overline{B} : \mathbb{KP}^n \to \mathbb{KP}^n \) are equal.

**Proof.** By definition of \( \operatorname{PGL}_n(\mathbb{K}) \), if \( A \) and \( B \) have the same image in \( \operatorname{PGL}_n(\mathbb{K}) \), then we can write \( A = \lambda B \) for some \( \lambda \in \mathbb{K}^* \) and we already noted above that this implies that the projective transformations \( \overline{A} \) and \( \overline{B} \) are equal.

Conversely, suppose the projective transformations \( \overline{A}, \overline{B} : \mathbb{KP}^n \to \mathbb{KP}^n \) are equal. Then, since \( \overline{A}[e_i] = \overline{B}[e_i] \) for \( i = 0, \ldots, n \), we can write \( A e_i = \lambda_i B e_i \) for some \( \lambda_i \in \mathbb{K}^* \). Also, since \( \overline{A}[1 : \cdots : 1] = \overline{B}[1 : \cdots : 1] \), we can write

\[
A e_0 + \cdots + A e_n = \lambda (B e_0 + \cdots + B e_n)
\]

for some \( \lambda \in \mathbb{K}^* \). Putting these together we see that

\[
\lambda_0 B e_0 + \cdots + \lambda_n B e_n = \lambda B e_0 + \cdots + \lambda B e_n.
\]

But \( B \) is invertible, so \( B e_0, \ldots, B e_n \) is a basis for \( \mathbb{K}^{n+1} \), hence the above equality implies that \( \lambda = \lambda_1 = \cdots = \lambda_n \) and we conclude that \( A = \lambda B \), so \( A \) and \( B \) have the same image in \( \operatorname{PGL}_n(\mathbb{K}) \).

**Theorem 4.4.4.** The subgroup

\[
G := \{ \overline{A} \in \operatorname{PGL}_{n+1}(\mathbb{K}) : \overline{A}(\mathbb{K}^n) = \mathbb{K}^n \}
\]

of \( \operatorname{PGL}_{n+1}(\mathbb{K}) \) consisting of projective transformations that “preserve the locus of finite points \( \mathbb{K}^n \subseteq \mathbb{KP}^n \)” is isomorphic to the group \( \operatorname{Aff}(\mathbb{K}^n) \) of affine transformations of \( \mathbb{K}^n \) via the restriction map \( \overline{A} \mapsto \overline{A}|_{\mathbb{K}^n} \).
Proof. Suppose $A \in G$, so the projective transformation $\overline{A} : \mathbb{P}^n \to \mathbb{P}^n$ takes the locus $\mathbb{K}^n \subseteq \mathbb{P}^n$ of finite points into itself. Equivalently, since $\overline{A} : \mathbb{P}^n \to \mathbb{P}^n$ is bijective, this means that $\overline{A}$ takes the locus $\mathbb{K}^{n-1} \subseteq \mathbb{P}^n$ of infinite points into itself. Pick $A \in \text{GL}_{n+1}(\mathbb{K})$ mapping to $\overline{A}$ in $\text{PGL}_n(\mathbb{K})$. Since $[e_0], \ldots, [e_{n-1}]$ are infinite points of $\mathbb{P}^n$, the points $\overline{A}[e_i] = [Ae_i]$ must also be infinite points for $i = 0, \ldots, n - 1$, which means that the last entry of each vector $Ae_i$ ($i = 0, \ldots, n - 1$) must be zero. On the other hand, $A$ must be invertible, so the last entry $\lambda$ of $Ae_n$ therefore cannot be zero. Then $L(\overline{A}) := \lambda^{-1} A \in \text{GL}_{n+1}(\mathbb{K})$ also maps to $\overline{A}$ in $\text{PGL}_n(\mathbb{K})$ and is the unique such matrix with lower right entry equal to 1. Note that the $(n+1) \times (n+1)$ matrix $L(\overline{A})$ takes the “block upper triangular” form

$$L(\overline{A}) = \begin{pmatrix} B & t \\ 0 & 1 \end{pmatrix},$$

where $B$ is an $n \times n$ matrix and $t$ is a column vector with $n$ rows. A matrix in this block upper triangular form is invertible iff $B$ is invertible (for example, because one can check that $\det L(\overline{A}) = \det B$ for such a “block upper triangular” matrix $L(\overline{A})$). From the definition of matrix multiplication we see that two such block upper triangular matrices multiply according to the rule

$$\begin{pmatrix} B & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} B' & t' \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} BB' & Bt' + t \\ 0 & 1 \end{pmatrix}.$$  \hspace{1cm} (4.3)

In particular, this shows that $L(\overline{A_1} \overline{A_2}) = L(\overline{A_1}) L(\overline{A_2})$, so that $L$ may be viewed as an isomorphism of groups from $G$ to the subgroup $G'$ of $\text{GL}_{n+1}(\mathbb{K})$ consisting of invertible matrices of the “block upper triangular form” described above. Furthermore, if we compare the multiplication formula (4.3) to the formula (1.3) for composition of affine transformations, we see that

$$\begin{pmatrix} B & t \\ 0 & 1 \end{pmatrix} \mapsto [B, t]$$

is an isomorphism from $G'$ to $\text{Aff}(\mathbb{K}^n)$. To see that the isomorphism $\overline{A} \mapsto [B, t]$ from $G$ to $\text{Aff}(\mathbb{K}^n)$ just described is given by the indicated restriction map, just notice that if we fix a point $\overline{x} \in \mathbb{K}^n$, then the corresponding point of $\mathbb{P}^n$ is $[\overline{x} : 1]$ and we have

$$\overline{A}[\overline{x} : 1] = [L(\overline{A})(\overline{x}, 1)] = \begin{pmatrix} B \overline{x} + t \end{pmatrix},$$

which is nothing but $[B, t](\overline{x}) = B\overline{x} + t \in \mathbb{K}^n$ viewed as a point of $\mathbb{P}^n$ in the usual manner. $\Box$
We now prove the analogue of Proposition 1.4.3 for projective transformations.

**Proposition 4.4.5.** If $P_0, \ldots, P_{n+1}$ are $n + 2$ points of $\mathbb{K}P^n$ in general position (Definition 4.2.2), then there is a unique $A \in \text{PGL}_n(\mathbb{K})$ satisfying $A[e_i] = P_i$ for $i = 0, \ldots, n$ and $A[1: \cdots : 1] = P_{n+1}$.

**Proof.** Choose representatives $x_0, \ldots, x_{n+1} \in \mathbb{K}^{n+1} \setminus \{0\}$ for the $P_i$. Since the $P_i$ are in general position, $x_0, \ldots, x_n$ must be a basis for $\mathbb{K}^{n+1}$, so we can write

$$x_{n+1} = \sum_{i=0}^{n} \lambda_i x_i$$

for unique $\lambda_i \in \mathbb{K}$. Each $\lambda_i$ must be non-zero, otherwise $x_{n+1}$ would be in the span of a proper subset of $\{x_0, \ldots, x_n\}$, in violation of the general position assumption. The image $\overline{A} \in \text{PGL}_{n+1}(\mathbb{K})$ of the matrix $A \in \text{GL}_{n+1}(\mathbb{K})$ whose $i^{th}$ column is $\lambda_i x_i$ is as desired. \(\square\)

### 4.5 Exercises

**Exercise 4.1.** In Definition 4.1.2 we defined subsets $U_0, U_1, U_2$ of $\mathbb{K}P^2$, and, more generally, $U_0, \ldots, U_n$ of $\mathbb{K}P^n$. In Definition 4.1.3 we defined bijections $\phi_i : \mathbb{K}^2 \to U_i \subseteq \mathbb{K}P^2$, and, more generally, $\phi_i : \mathbb{K}^n \to U_i \subseteq \mathbb{K}P^n$.

a) Describe the subsets $\phi_1^{-1}(U_1 \cap U_2)$, $\phi_1^{-1}(U_0 \cap U_1 \cap U_2)$, and $\phi_2^{-1}(U_1 \cap U_2)$ of $\mathbb{K}^2$.

b) We obtain a bijection $\phi_{12}$ from $\phi_1^{-1}(U_1 \cap U_2)$ to $\phi_2^{-1}(U_1 \cap U_2)$ by mapping a point in $\phi_1^{-1}(U_1 \cap U_2)$ into $U_1 \cap U_2$ using the map $\phi_1$, then taking the inverse of the resulting point under $\phi_2$. Give an explicit formula for $\phi_{12}$.

c) More generally, for an arbitrary projective space $\mathbb{K}P^n$, describe the subsets $\phi_i^{-1}(U_i \cap U_j)$ and the bijections $\phi_{ij} : \phi_i^{-1}(U_i \cap U_j) \to \phi_j^{-1}(U_i \cap U_j)$ defined in the analogous manner.

The next three exercises probably require some background in topology.

**Exercise 4.2.** Show that, for a subset $U \subseteq \mathbb{R}P^n$, the conditions of Definition 4.3.1 are equivalent. Also show that the conditions of Definition 4.3.2 are equivalent for a subset $U \subseteq \mathbb{C}P^n$.

**Exercise 4.3.** For $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, prove that the subsets $U_i \subseteq \mathbb{K}P^n$ defined in Definition 4.1.2 are open in the topology on $\mathbb{K}P^n$ defined in Definitions 4.3.1 and 4.3.2. (If you haven’t done the previous exercise, then just use, say, the first condition in each of those definitions.) Prove that the bijections $\phi_i : \mathbb{K}^n \to U_i$ of Definition 4.1.3 are homeomorphisms when $U_i$ is given the subspace topology from the inclusion $U_i \subseteq \mathbb{K}P^n$.

**Exercise 4.4.** Show that $\mathbb{C}P^1$ is homeomorphic to the 2-sphere $S^2$.

**Exercise 4.5.** Let $A \in \text{GL}_{n+1}(\mathbb{K})$, $x \in \mathbb{K}^{n+1} \setminus \{0\}$. In terms of basic concepts from linear algebra, what does it mean to say that $[x] \in \mathbb{K}P^n$ is a fixed point of the projective transformation $\overline{A} : \mathbb{K}P^n \to \mathbb{R}P^n$ associated to $A$?
Exercise 4.6. Let $L_1, L_2, L_3$ be three lines in $\mathbb{R}^2$ with empty intersection: $L_1 \cap L_2 \cap L_3 = \emptyset$. How many connected components does $\mathbb{R}^2 \setminus (L_1 \cup L_2 \cup L_3)$ have? What happens if $\mathbb{R}^2$ is replaced by $\mathbb{RP}^2$? \textit{Hint:} Move one of the lines to infinity by a projective transformation. (In fact you can move them even “more freely” as described in Exercise 4.10.)

There are an infinite number of possible variations on Proposition 4.4.5. Here are some:

Exercise 4.7. Suppose $L$ and $M$ are lines in $\mathbb{KP}^3$ and $P$ is a point of $\mathbb{KP}^3$. Prove that there is a line in $\mathbb{KP}^3$ containing $P$ and intersecting both $L$ and $M$.

Exercise 4.8. Suppose $\mathbb{P}(V) \subseteq \mathbb{KP}^n$ is a linear subspace. Prove that every projective transformation of $\mathbb{P}(V)$ (invertible linear transformation $V \rightarrow V$, up to rescaling) extends to a projective transformation of $\mathbb{KP}^n$.

Exercise 4.9. Let $L$ be a line in $\mathbb{KP}^2$, $P$ a point of $L$, $M$ a line not containing $P$. Prove that there is a projective transformation $\overline{A}: \mathbb{KP}^2 \rightarrow \mathbb{KP}^2$ taking $L$ to the line at infinity, $P$ to the point $[0 : 1 : 0]$, and $M$ to the line $Z(y) = \mathbb{P}(\text{Span}(e_0, e_1))$.

Exercise 4.10. Suppose $L_1, \ldots, L_4$ are four lines in $\mathbb{KP}^2$ such that the intersection of any three is empty. Suppose $L'_1, \ldots, L'_4$ are four lines in $\mathbb{KP}^2$ with the same property. Prove that there is a unique projective transformation $\overline{A}: \mathbb{KP}^2 \rightarrow \mathbb{KP}^2$ taking $L_i$ onto $L'_i$ for $i = 1, \ldots, 4$. \textit{Hint:} This statement is “dual” to the statement of Proposition 4.4.5 in the case $n = 2$. 
Chapter 5

Projective Plane Curves

In this chapter, we will specialize our study of projective spaces in Chapter 4 to the case of the projective plane $\mathbb{P}^2$ over a field $\mathbb{K}$. Recall that we identify the affine plane $\mathbb{A}^2$ over $\mathbb{K}$ with the subset $U_2 = \{(x : y : z) \in \mathbb{P}^2 : z \neq 0\}$ of the projective plane $\mathbb{P}^2$ via the map $(x, y) \mapsto [x : y : 1]$, thus we view $\mathbb{A}^2$ as a subset of $\mathbb{P}^2$, which we call the set of finite points. The complementary subset of infinite points is called the line at infinity; it is identified with $\mathbb{P}^1$ via $[x : y : 0] \leftrightarrow [x : y]$. We will see that the projective plane has many nice features that are lacking in the affine plane $\mathbb{A}^2$. For example, in $\mathbb{A}^2$, two distinct lines may or may not intersect, whereas in $\mathbb{P}^2$, any two distinct lines intersect in precisely one point ($\S 5.4$).

Just as a polynomial in two variables $f \in \mathbb{K}[x, y]$ gives rise to an affine algebraic plane curve $Z(f) \subseteq \mathbb{A}^2$, we shall see in $\S 5.1$ that a homogeneous polynomial (also called a form) in three variables $f \in \mathbb{K}[x, y, z]$ gives rise to a projective plane curve $Z(f) \subseteq \mathbb{P}^2$. Using this construction, we shall revisit some of the geometry we studied in Chapter 3.

In $\S 5.8$ we give a brief treatment of intersection multiplicity and Bezout’s Theorem, to the extent that we shall need it elsewhere. This section is rather tedious and not a particularly satisfactory treatment of intersection multiplicity, as my hands were tied by my refusal to make use of any concepts from commutative algebra (besides the notion of fields and polynomials). In $\S 5.5$ we see that the discriminant of a real affine conic, introduced in $\S 3.6$ has a natural “geometric” interpretation in projective geometry. In $\S 5.6$ we carry out the classification of projective conics (over $\mathbb{R}$ and $\mathbb{C}$) promised in $\S 3.5$.

The remaining sections of this chapter are largely independent of one another and consist of a hodge-podge of topics in the geometry of projective plane curves. In $\S 5.7$, we discuss the cross ratio, which may be thought of as a classification of ordered lists of four points in $\mathbb{P}^1$, up to projective equivalence. In $\S 5.9$ we prove some “classical” theorems of projective geometry. The theorem of Desargues proved there plays an important role in axiomatic projective geometry. In $\S 5.10$ we show that the set of non-singular points of (certain) smooth cubic curves can be given the structure of an abelian group. We describe this group explicitly in the case of the “cuspital cubic” in $\S 5.11$. 

55
5.1 Homogeneous polynomials and zero loci

A polynomial \( f(x, y, z) \in \mathbb{K}[x, y, z] \) is called \textit{homogeneous of degree} \( d \) (or a \textit{form of degree} \( d \)) iff it can be written in the form

\[
 f(x, y, z) = \sum_{i+j+k=d} A_{ijk} x^i y^j z^k
\]

for coefficients \( A_{ijk} \in \mathbb{K} \). Usually we assume that at least one of the \( A_{ijk} \) is non-zero, so that \( f \neq 0 \), though in some situations it is convenient to have the convention that the zero polynomial is homogeneous of degree \( d \) for every \( d \). A form of degree one (resp. two, three, \ldots) is often called a \textit{linear} (resp. \textit{quadratic}, \textit{cubic}, \ldots) \textit{form}.

If \( f \) is a form of degree \( d \) then for any \( \lambda \in \mathbb{K} \) we have

\[
 f(\lambda x, \lambda y, \lambda z) = \sum_{i+j+k=d} (\lambda x)^i (\lambda y)^j (\lambda z)^k = \sum_{i+j+k=d} \lambda^{i+j+k} x^i y^j z^k = \lambda^d f(x, y, z).
\]

To \textit{any} degree \( d \) polynomial

\[
 f(x, y) = \sum_{i+j \leq d} A_{ij} x^i y^j
\]

in \textit{two variables} we associate a \textit{homogeneous polynomial} \( \tilde{f} \) in \textit{three variables}, called the \textit{homogenization} of \( f(x, y) \), by setting

\[
 \tilde{f}(x, y, z) := \sum_{i+j \leq d} A_{ij} x^i y^j z^{d-i-j}.
\]

This association \( f \mapsto \tilde{f} \) yields a bijection between degree \( d \) polynomials in \( x, y \) and homogeneous degree \( d \) polynomials in \( x, y, z \) whose inverse is given by \( f(x, y, z) \mapsto f(x, y, 1) \).

For any \textit{homogeneous} polynomial \( f(x, y, z) \), we can consider the subset

\[
 Z(f) := \{ [x : y : z] \in \mathbb{K}\mathbb{P}^2 : f(x, y, z) = 0 \}
\]

of the projective plane \( \mathbb{K}\mathbb{P}^2 \). This makes sense because the question of whether \( f(x, y, z) = 0 \) depends only on the \textit{equivalence class} \( [x : y : z] \) of \( (x, y, z) \in \mathbb{K}^3 \setminus \{0\} \), in light of the formula (5.1) (in the case of a non-zero \( \lambda \)).

\textbf{Definition 5.1.1.} For a homogeneous polynomial (form) \( f \in \mathbb{K}[x, y, z] \) (usually assumed to be of positive degree), the subset \( Z(f) \subseteq \mathbb{K}\mathbb{P}^2 \) defined by (5.2) is called the \textit{zero locus} of \( f \). A subset \( S \subseteq \mathbb{K}\mathbb{P}^2 \) equal to \( Z(f) \) for some homogeneous polynomial \( f \) (of positive degree) is called a \textit{projective plane curve} and \( f \) is called a \textit{defining equation} for \( S \). If \( f \in \mathbb{K}[x, y] \) is any non-constant polynomial, the zero locus \( Z(\tilde{f}) \subseteq \mathbb{K}\mathbb{P}^2 \) of the homogenization of \( f \) is called the \textit{projective completion} of \( Z(f) \subseteq \mathbb{K}^2 \).
5.2 Smooth and singular points

The usual remarks are in order: What we call the “projective completion of $Z(f)$” really depends highly on the polynomial $f$ defining $Z(f)$—it is not really intrinsic to the subset $Z(f) \subseteq \mathbb{K}^2$. Notice that, for a polynomial $f \in \mathbb{K}[x,y]$, we have

$$\tilde{f}(x,y,1) = f(x,y),$$

so $(x,y) \in \mathbb{K}^2$ is in $Z(f) \subseteq \mathbb{K}^2$ iff $[x:y:1] \in \mathbb{K}\mathbb{P}^2$ is in $Z(\tilde{f}) \subseteq \mathbb{K}\mathbb{P}^2$. In other words, the finite points of the projective completion $Z(\tilde{f}) \subseteq \mathbb{K}\mathbb{P}^2$ of $Z(f) \subseteq \mathbb{K}^2$ are just the points of $Z(f)$:

$$Z(\tilde{f}) \cap \mathbb{K}^2 = Z(f).$$

If $f \in \mathbb{K}[x,y,z]$ is homogeneous, notice that $f(x,y,0) \in \mathbb{K}[x,y]$ is homogeneous (though possibly zero even if $f$ is non-zero)—the intersection $Z(f) \cap \mathbb{K}\mathbb{P}^1$ of $Z(f) \subseteq \mathbb{K}^2$ with the line at infinity $\mathbb{K}\mathbb{P}^1 \subseteq \mathbb{K}\mathbb{P}^2$ is given by

$$Z(f) \cap \mathbb{K}\mathbb{P}^1 = \{[x:y] \in \mathbb{K}\mathbb{P}^1 : f(x,y,0) = 0\}.$$

This intersection with the line at infinity is often of interest even in the case where $f = \tilde{g}$ is the homogenization of a polynomial $g \in \mathbb{K}[x,y]$—it encodes, in an appealing geometric way, the “asymptotic” behaviour of $g(x,y)$ for “large” $(x,y) \in \mathbb{K}^2$. We shall return to this point in §5.5.

**Remark 5.1.2.** Given a homogeneous polynomial $f \in \mathbb{K}[x,y,z]$, the notation $Z(f)$, used in isolation could possibly be confusing, because it might be understood as the projective plane curve $Z(f) \subseteq \mathbb{K}\mathbb{P}^2$, or just as the “affine” zero locus

$$Z(f) = \{(x,y,z) \in \mathbb{K}^3 : f(x,y,z) = 0\}$$

of $f$. In these notes, we always have the first meaning of $Z(f)$ in mind. These two “zero loci” are not unrelated: The homogeneity of $f$ and (5.1) show that the “affine” zero locus of $f$—let us call it $C(Z(f))$ for the moment, for clarity—is “invariant under rescaling,” meaning: If $(x,y,z) \in C(Z(f))$, then $\lambda(x,y,z) \in C(Z(f))$ for any $\lambda \in \mathbb{K}$ (and conversely if $\lambda \in \mathbb{K}^\ast$). Note also that the projection map $\mathbb{K}^3 \setminus \{0\} \to \mathbb{K}\mathbb{P}^2$ given by $x \mapsto [x]$ takes $C(Z(f))$ (minus the origin) onto $Z(f) \subseteq \mathbb{K}\mathbb{P}^2$. For these reasons, the affine zero locus $C(Z(f))$ is often called the cone on the projective zero locus $Z(f) \subseteq \mathbb{K}\mathbb{P}^2$.

### 5.2 Smooth and singular points

In this section we give the projective analogues of the definitions from §3.2.

**Definition 5.2.1.** For a homogeneous polynomial $f \in \mathbb{K}[x,y,z]$, a point $P \in Z(f) \subseteq \mathbb{K}\mathbb{P}^2$ is called a non-singular (or smooth) point of $Z(f)$ iff at least one of $f_x(P)$, $f_y(P)$, $f_z(P)$ is non-zero. A point $P \in Z(f)$ for which $f_x(P) = f_y(P) = f_z(P) = 0$ is called a singular point of $Z(f)$. The subset of $Z(f)$ consisting of smooth (resp. singular) points is denoted $Z(f)^{\text{sm}}$ (resp. $Z(f)^{\text{sing}}$). We say that $f$ is non-singular or smooth iff $Z(f)^{\text{sing}}(\overline{\mathbb{K}})$ is empty for every field extension $\mathbb{K} \subseteq \overline{\mathbb{K}}$.\footnote{It equivalent to ask that $Z(f)^{\text{sing}}(\overline{\mathbb{K}}) = \emptyset$ for some algebraically closed field $\overline{\mathbb{K}}$ containing $\mathbb{K}$.}
As in the affine case, the question of whether \( P \in \mathbb{Z}(f) \) is a smooth point or a singular point depends highly on \( f \), not just on the curve \( C = \mathbb{Z}(f) \). The following lemma is often useful when trying to determine the singular points of a projective curve:

**Lemma 5.2.2.** For a homogeneous polynomial \( f \in \mathbb{K}[x, y, z] \) of degree \( d \), we have

\[
df = xf_x + yf_y + zf_z.
\]

Consequently, if \( d \) is not zero in \( \mathbb{K} \) (which holds if \( d > 0 \) and \( \mathbb{K} = \mathbb{Q}, \mathbb{R}, \mathbb{C} \), or any other field of characteristic zero), then a point \( P \in \mathbb{K}P^2 \) is a singular point of \( \mathbb{Z}(f) \) iff \( f_x(P) = f_y(P) = f_z(P) = 0 \).

**Proof.** Exercise 5.11.

**Definition 5.2.3.** If \( P \) is a non-singular point of \( \mathbb{Z}(f) \), we define the tangent line to \( \mathbb{Z}(f) \) at \( P \) to be the line \( T_P \mathbb{Z}(f) := \mathbb{Z}(xf_x(P) + yf_y(P) + zf_z(P)) \).

The previous lemma implies that \( P \in T_P \mathbb{Z}(f) \). The projective notions of smoothness, tangent lines, etc. are compatible in every imaginable sense with those from the affine setting of §3.2. Some precise statements along these lines are given in Exercise 5.12.

### 5.3 Linear change of variables and projective equivalence

The notion of *affine equivalence* introduced in §3.3 also has an analogue in the projective setting.

Given a form (homogeneous polynomial) \( f \in \mathbb{K}[x, y, z] = \mathbb{K}[\mathbf{x}] \) of degree \( d \) and an invertible matrix \( A \in \text{GL}_3(\mathbb{K}) \), we obtain a form \( f \cdot A \in \mathbb{K}[x] \) of degree \( d \) by setting

\[
(f \cdot A)(\mathbf{x}) := f(A\mathbf{x}).
\]

As in §3.3, we think of the vector \( \mathbf{x} = (x, y, z) \) of variables as a column vector, and the arguments of our polynomials as column vectors.

**Definition 5.3.1.** Two homogeneous polynomials \( f, g \in \mathbb{K}[x, y, z] = \mathbb{K}[\mathbf{x}] \) are called *projectively equivalent* iff there is an invertible matrix \( A \in \text{GL}_3(\mathbb{K}) \) and a \( \lambda \in \mathbb{K}^* \) so that \( f(\mathbf{x}) = \lambda g(A\mathbf{x}) \). Two subsets \( S, T \subseteq \mathbb{K}P^2 \) are called *projectively equivalent* iff there is an \( A \in \text{GL}_3(\mathbb{K}) \) such that \( A(S) = T \).

When \( g = f \cdot A \) for an invertible matrix \( A \), we often say that \( g \) is obtained from \( f \) by a *linear change of variables*. Similarly, if \( g = \lambda f \) for \( \lambda \in \mathbb{K}^* \), we say that \( g \) is a *rescaling* of \( f \).

The formula (5.1) shows that if \( A \in \text{GL}_3(\mathbb{K}) \) and \( f \in \mathbb{K}[x, y, z] \) is homogeneous of degree \( d \), then \( f \cdot (\lambda A) = \lambda^d f \cdot A \). So as long as we consider our homogeneous polynomials only up to rescaling (as is reasonable if we are concerned with their zero loci), \( f \cdot A \) depends only on the image \( A \in \text{PGL}_2(\mathbb{K}) \) of \( A \) in the projective general linear group.
As in the affine setting, it is straightforward to check that projective equivalence is in fact an equivalence relation (Exercise 5.1). As in the affine case (Lemma 3.3.1), one sees that when \( f(\overline{x}) = \lambda g(A\overline{x}) \), the corresponding projective plane curves are related by

\[
Z(f) = A^{-1}(Z(g)). \tag{5.3}
\]

In particular, projectively equivalent homogeneous polynomials have projectively equivalent zero loci. As in the affine case, it is easy to classify homogeneous linear polynomials up to projective equivalence—we shall see in the next section that there is only one equivalence class, regardless of the field \( \mathbb{K} \) over which we work.

### 5.4 Lines and linear functionals

Recall that, in Definition 4.2.1, we defined a line to be a subset \( L \) of \( \mathbb{K}P^2 \) equal to \( \mathbb{P}(V) \) for some two dimensional subspace \( V \) of the \( \mathbb{K} \) vector space \( \mathbb{K}^3 \). By linear algebra, every such \( V \) arises as the kernel of some surjective (equivalently: non-zero) linear transformation \((A, B, C) : \mathbb{K}^3 \to \mathbb{K} \), hence a line may equivalently be defined as a subset \( L \subseteq \mathbb{K}P^2 \) such that

\[
L = \{(x : y : z) \in \mathbb{K}P^2 : Ax + By + Cz = 0\}
\]

for some \((A, B, C) \neq (0, 0, 0)\) in \( \mathbb{K}^3 \). In other words, a line is the zero locus in \( \mathbb{K}P^2 \) of a linear form. As in Definition 5.1.1, we call \("Ax + By + Cz\) a defining equation for \( L \). (Compare the affine version of a line, discussed in §3.3.)

Recall from linear algebra that if \( V \) is a \( \mathbb{K} \)-vector space, then a linear transformation \( V \to \mathbb{K} \) is called a linear functional on \( V \). The set

\[
V^* := \text{Hom}_\mathbb{K}(V, \mathbb{K})
\]

of linear functionals on \( V \) is itself a \( \mathbb{K} \)-vector space (under "pointwise addition and scalar multiplication"), called the dual vector space of \( V \). The next lemma says that \([f] \mapsto \mathbb{P}(\text{Ker } f)\) establishes a bijection from \( \mathbb{K}P^{2*} = \mathbb{P}(\mathbb{K}^3)^* \) to the set of lines in \( \mathbb{K}P^2 \).

**Lemma 5.4.1.** The line with defining equation \( Ax + By + Cz \) coincides with the line with defining equation \( A'x + B'y + C'z \) iff \((A, B, C) = \lambda(A', B', C')\) for some \( \lambda \in \mathbb{K}^* \). In other words, the kernel of a linear functional \( f : \mathbb{K}^3 \to \mathbb{K} \) coincides with the kernel of the linear functional \( f' : \mathbb{K}^3 \to \mathbb{K} \) iff \( f = \lambda f' \) for some \( \lambda \in \mathbb{K}^* \).

**Proof.** If \((A', B', C') = \lambda(A, B, C)\) for some \( \lambda \in \mathbb{R}^* \), then \( Z(Ax + By + Cz) = Z(A'x + B'y + C'z) \) because \( A'x + B'y + C'z = \lambda(Ax + By + Cz) \), so \( A'x + B'y + C'z = 0 \) iff \( Ax + By + Cz = 0 \). On the other hand, suppose that \( Z(Ax + By + Cz) = Z(A'x + B'y + C'z) \) and let us show that \((A', B', C') = \lambda(A, B, C)\) for some \( \lambda \in \mathbb{R}^* \). After possibly permuting the three coordinates everywhere, we can assume that \( A \neq 0 \). Then \([-B : A : 0] \in Z(Ax + By + Cz) \subseteq \mathbb{R}P^2 \), so, since \( Z(Ax + By + Cz) = Z(A'x + B'y + C'z) \), we have \([-B : A : 0] \in Z(A'x + B'y + C'z)\), hence \( AB' = BA' \), or, equivalently, \( B' = \lambda B \), where we set \( \lambda := A'/A \). Similarly, since \([-C : 0 : A] \in Z(Ax + By + Cz) \), we find that \( AC' = CA' \), hence \( C' = \lambda C \), and then \((A', B', C') = \lambda(A, B, C)\) as desired. \( \square \)
In other words, lines in $\mathbb{KP}^2$ are in bijective correspondence with points of “another” projective plane $\mathbb{KP}^{2*}$ (whose coordinates we will call $A, B, C$, rather than $x, y, z$ to avoid confusion) via the map taking the line with defining equation $Ax + By + Cz$ to the point $[A : B : C]$. The following result sums up basically everything there is to know about lines:

**Proposition 5.4.2.** Let $\mathbb{K}$ be a field, $\mathbb{KP}^2$ the projective plane over $\mathbb{K}$.

1. For any line $L = Z(Ax + By + Cz)$, any point $P$ of $L$ is a smooth point and $L = T_P L$ is the tangent line to $L$ at $P$.

2. The “line at infinity” $\mathbb{KP}^1 \subseteq \mathbb{KP}^2$ is a line.

3. Any line $L$, other than the line at infinity, is the projective completion of a unique line in $\mathbb{K}^2$.

4. Any two distinct lines in $\mathbb{KP}^2$ intersect in exactly one point.

5. If $L \subseteq \mathbb{KP}^2$ is a line and $A \in GL_3(\mathbb{K})$, then the image $\overline{A}(L) \subseteq \mathbb{KP}^2$ of $L$ under the projective transformation $\overline{A} : \mathbb{KP}^2 \to \mathbb{KP}^2$ is also a line.

6. For any line $L$, there is some $A \in GL_3(\mathbb{K})$ such that $\overline{A}(L)$ is the line at infinity.

**Proof.** (1): The partial derivations of $f = Ax + By + Cz$ are $f_x = A$, $f_y = B$, and $f_z = C$, at least one of which is a non-zero element of $\mathbb{K}$ by definition of a line. The equality $L = T_P L$ is immediate from the definition of $T_P L$ (Definition 3.2.2).

(2): Note that $z = 0x + 0y + 1z$ is a defining equation for the line at infinity. Alternatively, note that the line at infinity is the projecitivization of the subspace of $\mathbb{K}^3$ spanned by the first two standard basis vectors $e_0, e_1$ (cf. §4.2).

(3): If $L = Z(Ax + By + Cz)$ is not the line at infinity, then by Lemma 5.4.1 and the description of the line at infinity as $Z(z)$, we see that either $A$ or $B$ is non-zero—let us assume $A \neq 0$ as the case $B \neq 0$ is similar. Then $Z(x + (B/A)y + (C/A))$ is a line in $\mathbb{K}^2$ whose projective completion is

$$Z(x + (B/A)y + (C/A)z) = Z(Ax + (B/A)y + (C/A)z) = L.$$ 

If $Z(A'x + B'y + C')$ (with $(A', B') \neq (0, 0)$) is another line in $\mathbb{K}^2$ whose projective completion is $L$, then we have $L = Z(A'x + B'y + C'z)$, hence $(A', B', C') = \lambda(A, B, C)$ for some $\lambda \in \mathbb{K}$ by Lemma 5.4.1. Then we have

$$(\lambda A)(x + (B/A)y + (C/A)z) = A'x + B'y + C',$$

and hence

$$Z(x + (B/A)y + (C/A)z) = Z(A'x + B'y + C').$$

(4): Suppose $L = Z(Ax + By + Cz)$ and $L' = Z(A'x + B'y + C'z)$ are two distinct lines. Then $(A, B, C), (A', B', C') \in \mathbb{K}^3$ are linearly independent by Lemma 5.4.1, so, by linear algebra, the matrix $A$ whose rows are $(A, B, C)$ and $(A', B', C')$ gives a surjective
The discriminant (Definition 3.6.1 in §3.6) of a degree two polynomial \( f(x, y) \in \mathbb{R}[x, y] \) has an elegant geometric interpretation in terms of projective plane curves, which we will give momentarily after a brief lemma.

**Lemma 5.5.1.** For \( A, B, C \in \mathbb{R} \) (not all zero), the subset

\[
S := \{ [x : y] \in \mathbb{RP}^1 : A x^2 + B x y + C y^2 = 0 \}
\]

of the real projective line is empty (resp. consists of a single one point, consists of two distinct points) if \( B^2 - 4 AC < 0 \) (resp. \( B^2 - 4 AC = 0, B^2 - 4 AC > 0 \)).

**Proof.** First of all, the point \( \infty = [1 : 0] \in \mathbb{RP}^1 \) is in \( S \) iff \( A = 0 \). Every point other than \( \infty \) in \( \mathbb{RP}^1 \) can be written as \([x : 1]\) for a unique real number \( x \), and such a point is in \( S \) iff \( A x^2 + B x + C = 0 \).

Now we divide into cases. If \( A = 0 \), then \( \infty \in S \) and \([x : 1]\) is in \( S \) iff \( B x + C = 0 \).

The equation \( B x + C = 0 \) has exactly one solution when \( B \neq 0 \) and no solutions when \( B = 0 \) (because then \( C \neq 0 \) since we assume \( A, B, C \) aren’t all zero). In the latter case, we have \( B^2 - 4 AC = B^2 > 0 \), and \( S \) has two points (namely \( \infty, [-C/B : 1] \)), while in the former case, we have \( B^2 - 4 AC = 0 \) and \( S \) consists only of the point \( \infty \). The lemma is hence correct in the case \( A = 0 \).

If \( A \neq 0 \), then \( \infty \notin S \), and hence the points of \( S \) are just the real numbers \( x \) for which \( A x^2 + B x + C = 0 \). Since \( A \neq 0 \), the number of such real numbers \( x \) is as stated in the lemma by a an exercise with the Quadratic Formula that should be familiar from high-school algebra. \( \square \)

**Remark 5.5.2.** Over the complex numbers, the corresponding result says that \( S \) consists of a single point if \( B^2 - 4 AC = 0 \) and two points otherwise. The “single” point should be thought of as occurring with “multiplicity two.” Cf. Bezout’s Theorem (§5.8).

Now we can see that the discriminant of a degree two polynomial \( f(x, y) \in \mathbb{R}[x, y] \) is just a way of encoding how the projective completion \( \mathbb{Z}(f) \subseteq \mathbb{RP}^2 \) of \( f \) intersects the line at infinity \( \mathbb{RP}^1 \subseteq \mathbb{RP}^2 \).

**Proposition 5.5.3.** Let \( f = A x^2 + B x y + C y^2 + D x + E y + F \) be a polynomial of degree at most two with real coefficients \( A, \ldots, F \) and discriminant \( D(f) = B^2 - 4 AC \). Then:

1. \( D(f) < 0 \) iff \( \mathbb{Z}(\tilde{f}) \) is disjoint from the line at infinity,
Chapter 5 Projective Plane Curves

2. \( D(f) > 0 \) iff \( Z(\tilde{f}) \cap \mathbb{RP}^1 \) consists of two points, and
3. \( D(f) = 0 \) iff \( Z(\tilde{f}) \cap \mathbb{RP}^1 \) consists of a single point.

Proof. Note that the homogenization of \( f \) is
\[
\tilde{f} = Ax^2 + Bxy + Cy^2 + Dxz + Eyz + Fz^2,
\]
so a typical point \([x : y : 0]\) of the line at infinity is on \( Z_\mathbb{P}(f) \) iff \( \tilde{f}(x, y, 0) = 0 \) iff
\[
Ax^2 + Bxy + Cy^2 = 0.
\]
The result now follows from the previous lemma.

For a generalization of these ideas, see Exercise 5.2.

5.6 Classification of projective conics

Just as we studied the problem of classifying affine plane curves and polynomials in two variables up to affine equivalence (§3.3), we can also study the problem of classifying projective plane curves and homogeneous polynomials in three variables up to projective equivalence (Definition 5.3.1). In this section, we will classify homogeneous degree two polynomials (with real or complex coefficients) and the corresponding projective plane curves (also called conics, as in the affine case) up to projective equivalence. The projective classification is simpler than the affine one, so we can give complete proofs in this section. As in the affine case, the classification of conics is highly sensitive to the field over which we work. The classification is “easier” over \( \mathbb{C} \) than it is over \( \mathbb{R} \).

Many steps in the classification make sense over an arbitrary field \( \mathbb{K} \) where \( 2 = 1+1 \in \mathbb{K}^* \) (i.e. a field whose characteristic is not equal to 2). Consider a quadratic form
\[
f = Ax^2 + Bxy + Cy^2 + Dxz + Eyz + Fz^2 \tag{5.4}
\]
with coefficients \( A, \ldots, F \) in such a field \( \mathbb{K} \). As in §3.5, we consider the symmetric \( 3 \times 3 \) matrix
\[
M(f) := \begin{pmatrix} A & B/2 & D/2 \\ B/2 & C & E/2 \\ D/2 & E/2 & F \end{pmatrix} \tag{5.5}
\]
with entries in \( \mathbb{K} \)—notice that we have to assume \( 2 \in \mathbb{K}^* \) to be able to divide by 2 in defining the off-diagonal entries of our matrix \( M(f) \). We have
\[
f(\overline{x}) = \overline{x}^t M(f) \overline{x}, \tag{5.6}
\]
where \( \overline{x} \) is our vector of variables \((x, y, z)\), viewed as a column vector as usual. Notice that (5.6) is a little more natural looking than the analogous formula (3.4) we encountered in the affine case.

In particular, notice that for \( A \in \text{GL}_3(\mathbb{K}) \), the formula (5.6) gives
\[
f(A \overline{x}) = (A \overline{x})^t M(f) A \overline{x}
= \overline{A}^t \overline{x}^t M(f) A \overline{x},
\]
thus we see that $M(f)$ and $M(f \cdot A)$ are related by
\[ M(f \cdot A) = A^t M(f) A. \] (5.7)

**Proposition 5.6.1.** Let $\mathbb{K}$ be a field. For a quadratic form $f \in \mathbb{K}[x, y, z]$, the following are equivalent:

1. We can write $f = gh$ for linear forms $g, h \in \mathbb{K}[x, y, z]$, hence $Z(f) = Z(g) \cup Z(h)$ is a union of “two” (possibly equal) lines.

2. $Z(f)$ contains a line.

These equivalent conditions imply condition (3) below. If $2 \in \mathbb{K}^*$, then the following two conditions are equivalent:

3. There is a singular point in $Z(f)$.

4. The matrix $M(f)$ is singular (not invertible).

**Proof.** Obviously (1) implies (2). Conversely, suppose $Z(f)$ contains a line $L$. By Proposition 5.4.2(6) there is an invertible matrix $A \in \text{GL}_2(\mathbb{K})$ such that $\overline{A}^{-1}(L)$ is the line at infinity, hence $Z(f \cdot A)$ contains the line at infinity by Formula 5.3. If $f \cdot A = gh$ factors as a product of two homogeneous linear polynomials, then $f = (g \cdot A^{-1})(h \cdot A^{-1})$ also has such a factorization. We thus reduce to proving that if $Z(f)$ contains the line at infinity, then $f$ factors as a product of two homogeneous linear polynomials. Suppose $Z(f)$ contains the line at infinity. Then $Z(f)$ contains the three points $[0 : 1 : 0]$, $[1 : 0 : 0]$, and $[1 : 1 : 1]$, so we have
\[ f(0, 1, 0) = f(1, 0, 0) = f(1, 1, 0) = 0. \]

If we write $f$ as in (5.4), then these equalities are equivalent to $A = B = C = 0$, in which case we have the desired factorization:
\[ f = z(Dx + Ey + Fz). \]

(1) implies (3): Since the intersection of two lines in $\mathbb{K}\mathbb{P}^2$ is non-empty by Proposition 5.4.2(4), this follows from Exercise 5.3.

(3) iff (4): Notice that for $\overline{x} \in \mathbb{K}^3$, we have
\[ 2M(f)\overline{x} = (f_x(\overline{x}), f_y(\overline{x}), f_z(\overline{x})) \] (5.8)
(up to a transpose). By linear algebra, the matrix $M(f)$ is singular iff $M(f)\overline{x} = 0$ for some $\overline{x} \in \mathbb{K}^3 \setminus \{0\}$. From (5.8), we see that this is equivalent to saying there is some $[\overline{x}] \in \mathbb{K}\mathbb{P}^2$ for which
\[ f_x(\overline{x}) = f_y(\overline{x}) = f_z(\overline{x}) = 0. \]
Since $2 \in \mathbb{K}^*$ this is equivalent to saying there is a singular point in $Z(f)$ by Lemma 5.2.2. \qed
**Definition 5.6.2.** A homogeneous degree two polynomial \( f \in \mathbb{K}[x, y, z] \) is called *degenerate* iff \( Z(f) \) contains a singular point; otherwise \( f \) is called *non-degenerate*.

**Theorem 5.6.3.** Suppose \( \mathbb{K} \) is a field with \( 2 \in \mathbb{K}^* \) and \( f \in \mathbb{K}[x, y, z] \) is a homogeneous polynomial of degree two. Then there is an invertible matrix \( M \in \text{GL}_3(\mathbb{K}) \) such that

\[
f \cdot M = Ax^2 + By^2 + Cz^2
\]

for some \( A, B, C \in \mathbb{K} \).

*Proof.* Start with an arbitrary such \( f \) as in (5.4). Our goal is to show that by making a finite sequence of linear changes of variables, we can eliminate the “cross terms” \( xy, xz, \) and \( yz \). Our procedure for doing this is a careful version of *completing the square* which sometimes goes by the name of *Lagrange’s reduction*.

**Step 1.** Our first step is to eliminate the cross terms \( Bxy \) and \( Dxz \) involving the first variable \( x \). If we already have \( B = D = 0 \), then we proceed immediately to Step 3 below. If \( A \neq 0 \), then proceed to Step 2. If \( A = 0 \), but \( C \neq 0 \), just exchange the variables \( x \) and \( y \), then proceed to Step 2. If \( A = C = 0 \), but \( B \neq 0 \), then our polynomial

\[
Bxy + Dxz + Ey + Fz^2
\]

is taken via the linear change of variables \( (x, y, z) \mapsto (x, x + y, z) \) to the polynomial

\[
Bx^2 + (B + E)xy + (D + E)xz + Ey + Fz^2,
\]

which has non-zero \( x^2 \) coefficient, so we can proceed to Step 2. In the remaining case where \( A = C = 0 \), but \( D \neq 0 \), we can make a similar change of variables \( (x, y, z) \mapsto (x, y, x + z) \), then proceed to Step 2.

**Step 2.** At this step, we are given a polynomial

\[
Ax^2 + Bxy + Cy^2 + Dxz + Ey + Fz^2
\]

with \( A \neq 0 \). By making the linear change of variables

\[
(x, y, z) \mapsto (x - \frac{B}{2A} y - \frac{C}{2A} z, y, z)
\]

(we need \( 2 \in \mathbb{K}^* \) here) our polynomial becomes

\[
Ax^2 + \left( C + \frac{B^2}{4A} \right) y^2 + \left( \frac{BD}{2A^2} - \frac{BD}{A} + E \right) yz + \left( \frac{D^2}{4A^2} - \frac{D^2}{2A} + F \right) z^2.
\]

The exact coefficients are not important, but note that there are no cross terms involving \( x \), so we can proceed to Step 3.

**Step 3.** At this step we are given a polynomial

\[
Ax^2 + Cy^2 + Ey + Fz^2
\]

with no cross terms involving \( x \), and we wish to eliminate the one remaining cross term \( Ey \) (without creating any new cross terms). Of course, if \( E = 0 \), then we are done. If
C \neq 0$, proceed to Step 4. If $C = 0$, but $F \neq 0$, then interchange $y$ and $z$ and proceed to Step 4. The remaining case is where $C = F = 0$, but $E \neq 0$. In this case, our polynomial

$$Ax^2 + Eyz$$

is taken by the invertible change of variables $(x, y, z) \mapsto (x, y, y + z)$ to the polynomial

$$Ax^2 + Ey^2 + Eyz,$$

which has non-zero $y^2$ coefficient, hence we can proceed to Step 4.

**Step 4.** At this step we are given a polynomial

$$Ax^2 + Cy^2 + Eyz + Fz^2$$

with $C \neq 0$. By making the invertible change of variables

$$y \mapsto y - \frac{E}{2C}z$$

fixing $x$ and $z$, our polynomial becomes

$$Ax^2 + Cy^2 + \left(-\frac{E^2}{4C} + F\right)z^2,$$

which is of the desired form. \hfill \square

**Corollary 5.6.4.** Every (non-zero) homogeneous degree two polynomial $f \in \mathbb{R}[x, y, z]$ is projectively equivalent (Definition 5.3.1) to exactly one of the following:

1. $x^2 + y^2 + z^2$
2. $x^2 + y^2 - z^2$
3. $x^2 + y^2$
4. $x^2 - y^2$
5. $x^2$.

**Proof.** By Theorem 5.6.3, $f$ is equivalent to a polynomial of the form

$$Ax^2 + By^2 + Cz^2,$$

for some $A, B, C \in \mathbb{R}$. For each of the coefficients $A, B, C$ which is non-zero, we can rescale the corresponding variable by the inverse of the square root of the absolute value of its coefficient, thus we see that $f$ will be equivalent to a polynomial of the same form where, now, $A, B, C \in A, B, C \in \{0, 1, -1\}$. Multiplying everything through by $(-1)$ if necessary, we can assume there are at least as many $+1$’s as $-1$’s. Permuting the variables if necessary, we can assume the $+1$’s come first, then the $-1$’s, then the zeros. Since we assume that the polynomial we started with was non-zero, the polynomial we get at the end is non-zero. The resulting list of possibilities is tabulated in the statement of the corollary.

Now, it still remains to show that no two polynomials on the list are equivalent. But even the corresponding projective plane curves
1. Empty
2. Circle
3. Point \([0 : 0 : 1]\)
4. Two intersecting lines
5. One “doubled” “line”

can be distinguished by elementary topological concerns, with the exception that the “line” is a copy of \(\mathbb{RP}^1\), which is also homeomorphic to the circle; but this can be distinguished from the other “circle” on the grounds that projective transformations take lines to lines and the circle contains three non-collinear points.

Corollary 5.6.5. Every (non-zero) homogeneous degree two polynomial \(f \in \mathbb{C}[x, y, z]\) is projectively equivalent (Definition 5.3.1) to exactly one of the following:

1. \(x^2 + y^2 + z^2\)
2. \(x^2 + y^2\)
3. \(x^2\).

Proof. This is proved in exactly the same way as the previous corollary, except, in \(\mathbb{C}\), we can also make a change of variables like \(x \mapsto \sqrt{-1}x\) to ensure that every such \(f\) is projectively equivalent to one of the form \(Ax^2 + By^2 + Cz^2\) with \(A, B, C \in \{0, 1\}\).

Remark 5.6.6. Corollary 5.6.5 is valid over any algebraically closed field \(K\) with \(2 \in K^*\), by essentially the same proof—the point is that, in such a field \(K\), every element \(a\) is a square since the polynomial \(x^2 - a \in K[x]\) must have a root.

Remark 5.6.7. Theorem 5.6.3 also makes sense in an arbitrary number of variables. The proof is basically the same: Suppose

\[
\sum_{i=1}^{n} A_i x_i^2 + \sum_{i<j} B_{ij} x_i x_j
\]

is a typical degree two homogeneous polynomial over a field \(K\) with \(2 \in K^*\). Suppose we have eliminated all the cross terms involving \(x_1, \ldots, x_{k-1}\) (i.e. the second sum is actually over \(k \geq i < j\)) and we want to further eliminate the ones involving \(x_k\). To do this, proceed as follows:

**Step 1.** We first ask whether \(A_l \neq 0\) for some \(l \in \{k, \ldots, n\}\). If so, then we exchange \(x_k\) and \(x_l\) if necessary (obviously this preserves the property that cross terms involving \(x_1, \ldots, x_{k-1}\) are zero) to assume that, in fact, \(A_k \neq 0\), then we proceed to Step 2. In the other case where \(A_k = A_{k+1} = \cdots = A_n = 0\), we find some \(l \in \{k+1, \ldots, n\}\) with
5.7 The cross ratio

$B_{kl} \neq 0$ (if there is no such $l$, then all the cross terms involving $x_k$ are already zero, so there is nothing to do). We then make the linear change of variables $x_l \mapsto x_k + x_l$ (fixing the other $x_i$). It is easy to see that this doesn’t create any cross terms involving $x_1, \ldots, x_{k-1}$ and that the new $x_k^2$ coefficient is $B_{kl} \neq 0$, hence we can proceed to Step 2.

**Step 2.** At this step, we have a homogeneous degree two polynomial as above, with no cross terms involving $x_1, \ldots, x_{k-1}$, with the further property that $A_k \neq 0$. The change of variables

$$x_k \mapsto x_k - \frac{B_{k,k+1}}{2A_k} - \frac{B_{k,k+2}}{2A_k} - \ldots - \frac{B_{k,n}}{2A_k}$$

(fixing the other $x_i$) then yields a polynomial with no cross terms involving $x_1, \ldots, x_k$.

Obviously repeating Steps 1 and 2 successively for $k = 0, 1, \ldots, n - 1$ eliminates all the cross terms.

**Remark 5.6.8.** If $K$ is of characteristic 2, one cannot eliminate the cross terms, even in the 2 variable case (Exercise 5.6).

5.7 The cross ratio

Recall from Proposition 4.4.5 that any three distinct points of $KP^1$ can be moved to $0 = [0 : 1], 1 = [1 : 1], \infty = [1 : 0]$ (respectively) by a unique projective transformation $A \in PGL_2(K)$. Now consider four distinct points $P_1, P_2, P_3, P_4 \in KP^1$. By the same proposition, there is a unique $A \in PGL_2(K)$ so that

$$(AP_1, AP_2, AP_3) = (0, 1, \infty),$$

thus we can associate the element

$$\overline{AP}_4 \in KP^1 \setminus \{0, 1, \infty\} = K \setminus \{0, 1\}$$

of $K$ to our configuration $(P_1, \ldots, P_4)$ of four distinct points. The question arises: Can we give an explicit formula for $\overline{AP}_4$ in terms of the homogeneous coordinates of $P_1, \ldots, P_4$?

Indeed, if $P_i = [x_i : y_i]$, then we claim that the cross ratio

$$\chi(P_1, P_2, P_3, P_4) := [(x_4y_1 - x_1y_4)(x_3y_2 - x_2y_3) : (x_2y_1 - x_1y_2)(x_3y_4 - x_4y_3)]$$

(5.9)

is the desired formula. The key calculation we need is:

**Lemma 5.7.1.** For $A \in GL_2(K)$, we have

$$\chi(AP_1, AP_2, AP_3, AP_4) = \chi(P_1, P_2, P_3, P_4).$$

**Proof.** If we write

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$AP_i = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} ax_i + by_i \\ cx_i + dy_i \end{pmatrix}.$$
then \( AP_i = [ax_i + by_i : cx_i + dy_i] \), and we make the following calculation:

\[
(ax_i + by_i)(cx_j + dy_j) - (ax_j + by_j)(cx_i + dy_i) = (ad - bc)(x_iy_j - x_jy_i).
\]

Thus we see that, replacing \( x_i \) by \( ax_i + by_i \) and \( y_i \) by \( cx_i + dy_i \) everywhere in (5.9) will only rescale both coordinates by \((ad - bc)^2\), which is non-zero since \( \det A = (ad - bc) \neq 0 \) (as \( A \) is invertible) hence the corresponding point of \( \mathbb{P}^1 \) is unchanged.

If we now specialize the formula (5.9) to the case \( P_1 = 0 = [0 : 1] \), \( P_2 = 1 = [1 : 1] \), \( P_3 = \infty = [1 : 0] \), \( P_4 = [x : 1] \), we find

\[
\chi(0, 1, \infty, x) = [(x \cdot 1 - 0 \cdot 1)(1 \cdot 1 - 1 \cdot 0) : (1 \cdot 1 - 0 \cdot 1)(1 \cdot x - 0)] \\
= [x : 1] \\
= x,
\]

which shows that our formula has the desired property.

Notice that the formula (5.9) makes sense as long as no three of the \( P_i \) coincide, so we can in fact define the cross ratio

\[
\chi(P_1, P_2, P_3, P_4) \in \mathbb{P}^1
\]

for any four points \( P_1, \ldots, P_4 \in \mathbb{P}^1 \), as long as no three of them are the same. In fact, we have:

1. \( \chi(P_1, P_2, P_3, P_4) = 0 \) iff \( P_1 = P_4 \) or \( P_2 = P_3 \)
2. \( \chi(P_1, P_2, P_3, P_4) = 1 \) iff \( P_1 = P_3 \) or \( P_2 = P_4 \)
3. \( \chi(P_1, P_2, P_3, P_4) = \infty \) iff \( P_1 = P_2 \) or \( P_3 = P_4 \)

Now suppose we have a line \( L \subseteq \mathbb{P}^2 \). Write \( L = \mathbb{P}(V) \) for a two-dimensional subspace \( V \subseteq \mathbb{K}^3 \). Choose an isomorphism of vector spaces \( f : V \rightarrow \mathbb{K}^2 \). Given four points \( P_1, \ldots, P_4 \in L \), no three the same, we can define their cross ratio, using the chosen isomorphism \( f \), by setting

\[
\chi_f(P_1, P_2, P_3, P_4) := \chi(f(P_1), f(P_2), f(P_3), f(P_4)) \in \mathbb{P}^1.
\]

We claim that this is independent of the choice of \( f \). Indeed, suppose \( g : V \rightarrow \mathbb{K}^2 \) is another vector space isomorphism. Then by linear algebra, we can write \( g = Af \) for a vector space isomorphism (invertible \( 2 \times 2 \) matrix) \( A : \mathbb{K}^2 \rightarrow \mathbb{K}^2 \). We then compute

\[
\chi_g(P_1, \ldots, P_4) = \chi(g(P_1), g(P_2), g(P_3), g(P_4)) \\
= \chi(AF(P_1), AF(P_2), AF(P_3), AF(P_4)) \\
= \chi(f(P_1), f(P_2), f(P_3), f(P_4)) \\
= \chi_f(P_1, P_2, P_3, P_4),
\]

using Lemma 5.7.1 for the second equality. Having established that our notion of cross ratio is independent of the choice of \( f \), we can unambiguously denote it \( \chi(P_1, P_2, P_3, P_4) \).
5.8 Intersection multiplicity

As it will be necessary at many points later in the notes, we are now forced to discuss the idea of intersection multiplicity. Suppose that $f_1$ and $f_2$ are (non-constant) polynomials in $\mathbb{K}[x, y]$ (for some field $\mathbb{K}$), or (non-constant) homogeneous polynomials in $\mathbb{K}[x, y, z]$. Let $C_1 := \text{Z}(f_1)$, $C_2 := \text{Z}(f_2)$ be the corresponding affine algebraic curves in $\mathbb{K}^2$, as in §3.1, or projective curves in $\mathbb{K}\mathbb{P}^2$ (§5.1), as appropriate. Assume that there are only finitely many points in $C_1 \cap C_2$ in $\mathbb{K}^2$ (or $\mathbb{K}\mathbb{P}^2$, as appropriate) for some (equivalently any) algebraic closure $\overline{\mathbb{K}}$ of $\mathbb{K}$. (For simplicity, the reader may wish to keep in mind the case where $\mathbb{K} = \overline{\mathbb{K}} = \mathbb{C}$.) Under these assumptions, one can assign a positive integer $m(f_1, f_2, P)$ to each point $P \in C_1 \cap C_2$, called the intersection multiplicity. (One has to assume the finiteness of $C_1 \cap C_2$ over $\mathbb{K}$, rather than just over $\mathbb{K}$, because it could happen, for example, that $f_1 = f_2$ and the common curve $C = \text{Z}(f_i)$ has only one point $P$ over $\mathbb{K}$, and in this situation there is no “reasonable” way to define $m(f_1, f_2, P)$.) Although $m(f_1, f_2, P)$ certainly depends on the defining equations $f_i$ for the curves $C_i$, one often abusively writes $m(C_1, C_2, P)$ instead of $m(f_1, f_2, P)$. One can then prove the following general result:

**Theorem 5.8.1. (Bezout’s Theorem)** Suppose $\mathbb{K}$ is an algebraically closed field and $f_1, f_2 \in \mathbb{K}[x, y, z]$ are homogeneous polynomials of degrees $d_1, d_2 > 0$. Let $C_i := \text{Z}(f_i) \subseteq \mathbb{K}\mathbb{P}^2$. Assume that $C_1 \cap C_2$ is finite. Then

$$\sum_{P \in C_1 \cap C_2} m(f_1, f_2, P) = d_1d_2.$$ 

The general definition of $m(f_1, f_2, P)$ involves more commutative algebra than I wish to assume for the purposes of these notes, so it will not be given here. We will content ourselves with defining $m(f_1, f_2, P)$ only in the case where one of the $f_i$, say $f_1$, is of degree one. We will also prove Bezout’s Theorem in this case. This special case will be sufficient for our later purposes.

Since we force ourselves to avoid any commutative algebra and to avoid the general language of algebraic geometry, our definition of $m(f_1, f_2, P)$ will be rather tedious, though elementary. I will relegate almost everything to the exercises. Our construction is based on the idea of multiplicity of roots for a polynomial in a single variable, which we shall now briefly review.

Let $f \in \mathbb{K}[x]$ be a polynomial of degree $d > 0$. It is a general fact of commutative algebra that one can find a field $\overline{\mathbb{K}}$ containing $\mathbb{K}$ as a subfield such that, in $\overline{\mathbb{K}}[x]$, one can factor $f$ as a product of linear polynomials—that is, one can write

$$f(x) = C \prod_{i=1}^{d} (x - \alpha_i) \quad (5.10)$$

with $C \in \overline{\mathbb{K}}^*$, $\alpha_1, \ldots, \alpha_d \in \overline{\mathbb{K}}$. (For example, if $\mathbb{K}$ is $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, then the Fundamental Theorem of Algebra says one can take $\overline{\mathbb{K}} = \mathbb{C}$. There is, in some sense, a “smallest” such $\overline{\mathbb{K}}$, called a splitting field for $f$.) It is clear that $\alpha \in \mathbb{K}$ is a root of $f$ (i.e. $f(\alpha) = 0$) iff $\alpha$
is one of the \( \alpha_i \). If \( \alpha \) is a root of \( f \), we define its \textit{multiplicity} \( m = m(\alpha, f) \) to be

\[
m(\alpha, f) := |\{ i \in \{1, \ldots, d\} : \alpha = \alpha_i \}|.
\]

(5.11)

By construction, \( m \in \{1, \ldots, d\} \). Of course, the formula (5.11) also makes sense when \( \alpha \) is not a root of \( f \), in which case \( m(\alpha, f) = 0 \). If the factorization (5.10) of \( f \) can be done in \( K \) (which holds, for example, if \( K \) is algebraically closed), then we clearly have \( \sum_{\alpha} m(f, \alpha) = d \), where the sum runs over the roots of \( f \) in \( K \). This simple fact will ultimately yield our special case of Bezout’s Theorem. The reader is asked to establish some basic properties of this notion of “multiplicity” for roots of polynomials in Exercise 5.7.

\textbf{Definition 5.8.2.} Let \( f \in K[x, y, z] \) be a form. A smooth point \( P \) of \( Z(f) \) is called an \textit{inflection point} of \( Z(f) \) iff the intersection multiplicity of the tangent line \( T_P Z(f) \) and \( Z(f) \) at \( P \) is greater than 2. (Cf. Exercise 5.14.)

\textbf{5.9 The theorems of Pascal, Pappus, and Desargues}

In this section we shall prove some classic theorems of projective geometry. The first of these is a famous theorem of Blaise Pascal (1623-1662), which he apparently proved at the age of 16. Unfortunately his own proof of the theorem was lost, though several different proofs have accumulated since 1639. H. S. M. Coxeter includes a proof of Pascal’s Theorem in at least three of his books \([C1, 3.35], [C3, 7.21], [C2, 9.23]\). In fact, \([C3, 1.7]\) also gives the “simple proof” due to van Yzeren’s proof \([vY]\). Pascal’s Theorem is also mentioned in \([R, p. 140]\), and appears as an exercise in \([H]\). There are several “proofs” of Pascal’s Theorem on the Wikipedia, none of which is particularly clear in my opinion. Many “modern” proofs of Pascal’s Theorem use Bezout’s Theorem, which is ridiculous overkill. It is much harder to formulate and prove Bezout’s Theorem than to just prove Pascal’s Theorem from first principles. The proof we give here is the simplest I could come up with, relying only on a tiny bit of linear algebra and high-school level algebraic manipulation.

\textbf{Theorem 5.9.1. (Pascal’s Theorem)} Suppose \( P_1, \ldots, P_6 \) are six distinct points contained in a non-degenerate conic \( C \subseteq \mathbb{RP}^2 \). Then the points \( X := P_1P_2 \cap P_4P_5 \), \( Y := P_2P_3 \cap P_4P_6 \), and \( Z := P_3P_4 \cap P_1P_6 \) are collinear.

\textbf{Proof.} It follows from properties of intersection multiplicity that no three points of a non-degenerate conic can be collinear, hence \( P_1, \ldots, P_6 \) are in general position. A projective transformation takes lines to lines and conics to conics, so, after applying a projective transformation, we can assume (by Proposition 4.4.5) that

\[
\begin{align*}
P_1 &= [1 : 0 : 0] \\
P_2 &= [0 : 1 : 0] \\
P_3 &= [0 : 0 : 1] \\
P_4 &= [1 : 1 : 1] \\
P_5 &= [a : b : c] \\
P_6 &= [u : v : w].
\end{align*}
\]
The lines $P_1P_2, P_1P_3, P_1P_4, P_2P_4, P_3P_4$ are described, respectively, by the equations

$$z, y, y - z, x, x - z, x - y,$$

so the fact that $P_5$ and $P_6$ are not on any of these lines is equivalent to the fact that the following numbers are non-zero:

$$c, b, b - c, a, a - c, a - b, w, v - w, u, u - w, u - v.$$

We will divide by these non-zero numbers freely in what follows.

Since $P_1 \in C$ (resp. $P_2 \in C, P_3 \in C$), the equation

$$Ax^2 + Bxy + Cy^2 + Dxz + Eyz + Fz^2$$

defining $C$ has $A = 0$ (resp. $C = 0, F = 0$). Then $P_4 \in C$ implies $E = -B - D$. We can’t have $B = 0$ (because then our conic would be degenerate), so we can assume, after rescaling that $B = 1$, that this equation takes the form $y(x - z) + Dz(x - y)$. Solving for $D$ using the fact that $P_5 \in C$ yields

$$D = \frac{b(c - a)}{c(a - b)}.$$

The fact that $P_6$ is on $C$ is then equivalent to:

$$v(u - w) + \frac{b(c - a)}{c(a - b)}w(u - v) = 0$$

Clearing the (non-zero!) denominator and expanding out, the $bcuvw$ term cancels and we find:

$$abuw - bcuw + acvw = acuv - bcuv + abvw. \quad (5.12)$$

The line $P_1P_2$ is given by $z = 0$, so $X = P_1P_2 \cap P_3P_4$ is given by the linear combination of $P_4 = [1 : 1 : 1]$ and $P_5 = [a : b : c]$ with vanishing third coordinate, namely

$$X = [a - c : b - c : 0].$$

By similar reasoning, we find

$$Y = [0 : bu - av : cu - aw]$$
$$Z = [v : v : w].$$

To show that $X, Y,$ and $Z$ are colinear, we need to show that the coordinates of $Y$ are a linear combination of those of $X$ and $Z$, so we need to show that

$$Y = X - \frac{a - c}{v}Z.$$

Expanding this out and clearing (non-zero!) denominators, we see that this is equivalent to:

$$[0 : bu - av : cu - aw] = [0 : v(a - b) : w(a - c)].$$
which in turn is equivalent to
\[ w(a - c)(bu - av) = (cu - aw)v(a - b). \] (5.13)

But if we expand out and cancel the term \(-a^2vw\) common to both sides of (5.13), we see that (5.13) is the same as (5.12).

There is also a variant of Pascal’s theorem for certain configurations of six points on a degenerate conic—it can be proved by the same technique of using a projective transformation to move the points into some convenient locations, then doing a little elementary algebra:

**Theorem 5.9.2. (Pappus’ Theorem)** Let \( L \) and \( L' \) be distinct lines, \( A, B, C \) three distinct points of \( L \) not equal to \( P := L \cap L' \). Let \( A', B', C' \) be three points of \( L' \) distinct from \( P \). Then the points \( X := AB' \cap A'B', Y := AC' \cap A'C', \) and \( Z := BC' \cap B'C' \) are collinear.

**Proof.** Exercise 5.16

The same general technique can also be used to prove:

**Theorem 5.9.3. (Desargues’ Theorem)** For points \( A, B, C, A', B', C' \in \mathbb{KP}^2 \) in general position, the following are equivalent:

1. The three lines \( AA', BB', \) and \( CC' \) have non-empty intersection.
2. The three points \( X := AB \cap A'B', Y := AC \cap A'C', \) and \( Z := BC \cap B'C' \) are collinear.

**Proof.** Exercise 5.17

**Remark 5.9.4.** Desargues’ Theorem does not hold without the general position assumption, even if the six points are in “general enough” position that the statement of the theorem makes sense. See Exercise 5.18.

### 5.10 Cubic curves and the group law

Up to this point, our discussion of projective plane curves has been limited to the cases of lines and conics—the zero loci of linear and quadratic forms. Now it is time to look at the zero locus of a homogeneous degree three polynomial—a cubic form. For technical reasons, and to ease some calculations and exposition, we will often assume in this section that \( 2 \in \mathbb{K}^* \), and occasionally, that \( 3 \in \mathbb{K}^* \). For reasons that we will discuss momentarily, we will also restrict our attention to a certain class of homogeneous degree three polynomials:

**Definition 5.10.1.** Let \( \mathbb{K} \) be a field with \( 2 \in \mathbb{K}^* \). A degree three polynomial \( g(x, y) \in \mathbb{K}[x, y] \) is said to be in Weierstrass form iff

\[ g(x, y) = y^2 - f(x) \] (5.14)
for some monic (leading coefficient one) degree three polynomial \( f(x) \in \mathbb{K}[x] \). A cubic form \( \tilde{g}(x, y, z) \in \mathbb{K}[x, y, z] \) is called a Weierstrass cubic (or is said to be in Weierstrass form) if it is equal to the homogenization of a degree three polynomial \( g(x, y) \in \mathbb{K}[x, y] \) in Weierstrass form. In other words, a Weierstrass cubic is a cubic form \( \tilde{g} \) that can be written

\[
\tilde{g}(x, y, z) = y^2z - \tilde{f}(x, z),
\]

where \( \tilde{f}(x, z) \in \mathbb{K}[x, z] \) is the homogeneous degree three polynomial obtained by homogenizing a monic degree three polynomial \( f(x) \in \mathbb{K}[x] \).

Here are the basic features of Weierstrass cubics:

**Proposition 5.10.2.** Assume \( 2 \in \mathbb{K}^* \). Let \( g(x, y) = y^2 - f(x) \in \mathbb{K}[x, y] \) be a degree three polynomial in Weierstrass form, with homogenization \( \tilde{g}(x, y, z) \in \mathbb{K}[x, y, z] \).

1. The singular points of \( Z(\tilde{g}) \) are the points of the form \( (\alpha, 0) = [\alpha : 0 : 1] \), where \( \alpha \) is a repeated root of \( f \). In particular, \( Z(\tilde{g}) \) has at most one singular point and if \( f \) has no repeated roots in an algebraic closure of \( \mathbb{K} \), then \( Z(\tilde{g}) \) is smooth.

2. The point \( \infty := [0 : 1 : 0] \) is the unique point of \( Z(\tilde{g}) \) on the line at infinity. It is an inflection point (Definition 5.8.2) of \( Z(\tilde{g}) \) with tangent line \( T_\infty Z(\tilde{g}) \) equal to the line at infinity.

3. The polynomial \( \tilde{g}(x, y, z) \) is invariant under the linear change of variables \( (x, y, z) \mapsto (x, -y, z) \), hence \( Z(\tilde{g}) \) is invariant under the involution \( [x : y : z] \mapsto [x : -y : z] \) of \( \mathbb{K}P^2 \).

4. The points of \( Z(\tilde{g}) \) fixed by the involution from the previous part are the point \( \infty \), together with the points of the form \( (0, \alpha) = [\alpha : 0 : 1] \) where \( \alpha \) is a root of \( f \).

5. \( \tilde{g} \) cannot be factored as a product of a linear form \( g_1 \) and a quadratic form \( g_2 \), even if the coefficients of \( g_1 \) and \( g_2 \) are allowed to be in a field extension \( \mathbb{K} \supseteq \mathbb{K} \).

**Proof.** (1) and (2): The partial derivatives of \( g \) are \( g_x = -f'(x) \) and \( g_y = 2y \). Since \( 2 \in \mathbb{K}^* \), a point \( (a, b) \in \mathbb{K}^2 \) satisfies

\[
g(a, b) = g_x(a, b) = g_y(a, b) = 0
\]

iff \( b = 0 \) and \( f(a) = f'(a) = 0 \)—these latter equalities are equivalent to saying that \( a \) is a repeated root of \( f \) (Exercise 5.7(1)). This shows that the only “finite” singular points of \( Z(\tilde{g}) \) (cf. Exercise 5.12) are those described in (1). Since \( f(x) \) is monic of degree three, we have \( \tilde{f}(x, 0) = x^3 \) and hence \( \tilde{g}(x, y, 0) = x^3 \), hence \( \infty = [0 : 1 : 0] \) is the unique point of \( Z(\tilde{g}) \) on the line at infinity and the line at infinity intersects \( Z(\tilde{g}) \) at \( \infty \) with multiplicity three. Next notice that \( \tilde{f}_z(x, z) \) is homogeneous of degree two, so certainly \( \tilde{f}_z(0, 0) = 0 \). Then, since \( \tilde{g}_z = y^2 - \tilde{f}_z \), we see that \( \tilde{g}_z(0, 1, 0) = 1 \neq 0 \), so \( \infty \) is a smooth point of \( Z(\tilde{g}) \).

(3): The invariance of \( \tilde{g} \) under the indicated change of variables is clear since \( \tilde{g} \) is actually a polynomial in \( x, y^2, \) and \( z \), so the invariance of \( Z(\tilde{g}) \) under the indicated projective transformation is just a special case of the formula (5.3).
(4): The involution in question takes the locus $\mathbb{K}^2 \subseteq \mathbb{K}P^2$ of finite points into itself via the involution $(x, y) \mapsto (x, -y)$. Clearly then, a finite point is fixed by this involution iff it is of the form $(\alpha, 0)$. Such a point is in $Z(g)$ iff $f(\alpha) = 0$. On the line at infinity, the involution is given by $[x : y : 0] \mapsto [x : -y : 0]$. This has exactly two fixed points: $[1 : 0 : 0]$ and $[0 : 1 : 0]$, only the second of which is actually in $Z(\tilde{g})$.

(5): Suppose we could write $\tilde{g} = g_1 g_2$ in $\mathbb{K}[x, y, z]$ for some field $\mathbb{K} \supseteq K$ for a linear form $g_1 \in \mathbb{K}[x, y, z]$ and a quadratic form $g_2 \in \mathbb{K}[x, y, z]$. Since the coefficient of $x^3$ is 1 in $\tilde{g}$, we could rescale $g_1$ and $g_2$ to arrange that they take the form
\[
g_1 = x + Gy + Hz \\
g_2 = x^2 + Bxy + Cy^2 + Dxy + Eyz + Fz^2
\]
for some $B, \ldots, H \in \mathbb{K}$. Since $\tilde{g}(x, y, 0) = x^3$, we find that $B, \ldots, H$ must satisfy
\[
G + B = 0 \quad (5.16) \\
CG = 0 \quad (5.17) \\
C + BG = 0. \quad (5.18)
\]
First suppose that $G = 0$. Then these equations imply that $B = C = 0$. But then $g_1 g_2$ looks like
\[
(x^2 + Dxz + Eyz + Fz^2)(x + Hz),
\]
which has no $y^2z$ term despite the fact that the coefficient of $y^2z$ in $\tilde{g}$ is 1. Next suppose that $G \neq 0$. Then (5.17) implies that $C = 0$ and (5.18) implies that $B = 0$, but then (5.16) doesn’t hold. We conclude that there can be no such factorization. 

Now we prove that any cubic polynomial with these same geometric features is projectively equivalent to a Weierstrass cubic:

**Theorem 5.10.3.** Suppose $h \in \mathbb{K}[x, y, z]$ is a cubic form for which there exist an inflection point $P \in Z(h)$ and a projective transformation $\overline{M}$ satisfying:

1. $\overline{M}$ is an involution, meaning: $\overline{M}^2 = \text{Id}$,
2. the fixed locus of $\overline{M}$ consists of the point $P$ and some line $L$ not containing $P$, and
3. $h \cdot M$ is a non-zero multiple of $h$ for some (equivalently any) $M \in \text{GL}_3(\mathbb{K})$ mapping to $\overline{M}$ in $\text{PGL}_2(\mathbb{K})$.

Then $2 \in \mathbb{K}^*$ and $h$ is projectively equivalent to a cubic $g = y^2 z - \tilde{f}(x, z)$ in Weierstrass form. If, furthermore, $3 \in \mathbb{K}^*$, then we can also arrange that the coefficient of $x^2$ in $f$ is zero.

**Proof.** By Exercise 4.9, we can find $A \in \text{GL}_3(\mathbb{K})$ such that the projective transformation $\overline{A}$ takes $P$ to $\infty = [0 : 1 : 0]$, the tangent line $T_P Z(h)$ to the line at infinity, and the line $L$ to the line $Z(y)$. Replacing $h$, $P$, and $\overline{M}$ with $h \cdot A^{-1}$, $\infty$, and $\overline{A} \overline{M} \overline{A}^{-1}$, respectively, we can assume that $P = \infty$, $T_P Z(h) = Z(z)$, and $L = Z(y)$.
5.10 Cubic curves and the group law

Since $\infty$ is an inflection point for $\mathbb{Z}(h)$ and $T_\infty \mathbb{Z}(h) = \mathbb{Z}(z)$, $\mathbb{Z}(h)$ must intersect $\mathbb{Z}(z)$ only at $\infty$ with multiplicity three, hence $h(x, y, 0)$ must be a non-zero multiple of $x^3$, so, after rescaling $h$, we can assume $h(x, y, 0) = -x^3$. This means

$$h = By^2z - x^3 - Ax^2z - Cxz^2 + Dy^2z - Ez^3 + Fxyz.$$ 

Pick a lift $M \in \text{GL}_3(\mathbb{K})$ of $\overline{M}$. Since $\overline{M}$ fixes $\infty = [e_1]$ and $\mathbb{Z}(y) = \mathbb{P}(\text{Span}(e_0, e_2))$, $e_1$ must be an eigenvector for $M$, say with eigenvalue $\lambda$, and $\text{Span}(e_0, e_2)$ must be an eigenspace for $M$, say with eigenvalue $\mu$. We can't have $\lambda = \mu$, for then $M$ would be $\lambda = \mu$ times the identity and $\overline{M} = \text{Id}$ would fix all of $\mathbb{K}\mathbb{P}^2$. Since $\overline{M}^2 = \text{Id}$, we must have $\lambda^2 = \mu^2$, hence $\lambda = -\mu$. (Since $\lambda \neq \mu$ this implies $2 \in \mathbb{K}^\ast$.) Replacing $M$ with $\mu^{-1}M$, we can assume $\mu = 1$, $\lambda = -1$, hence $M : \mathbb{K}^3 \to \mathbb{K}^3$ is given by $(x, y, z) \mapsto (x, -y, z)$ and $\overline{M}$ is given by $[x : y : z] \mapsto [x : -y : z]$.

Since $h \cdot M$ must be a non-zero scalar multiple of $h$, we find that $D = F = 0$. We can't have $B = 0$, for then

$$h = -x^3 - Ax^2z - Cxz^2 - Ez^3$$

would be singular at $\infty$. So, after doing the change of variables $z \mapsto B^{-1}z$, we can assume $B = 1$, which puts $h$ in Weierstrass form:

$$h = y^2z - x^3 - A'x^2z - B'xz^2 - C'z^3.$$ 

If $3 \in \mathbb{K}^\ast$ we can do the linear change of variables $x \mapsto (x - A'/3z)$ to get rid of the $x^2z$ term. 

**Remark 5.10.4.** There are "better" reasons than Theorem 5.10.3 for restricting our attention to Weierstrass cubics. It turns out that any smooth cubic $C$ with a point $P \in C$ is "abstractly isomorphic" to a Weierstrass cubic by an "isomorphism" taking $P$ to $\infty$. In these notes, we have not developed the notion of an "abstract curve" or an "isomorphism" between such things—we have only the notion of projective equivalence. Now, it is not true that any smooth cubic $C$ with a point $P \in C$ is projectively equivalent to a Weierstrass cubic by a projective equivalence taking $P$ to $\infty$. This is simply because $P$ might not be an inflection point of $C$. Even "worse," there are smooth cubics $C$ (over $\mathbb{K} = \mathbb{Q}$, say) with non-empty zero locus (over $\mathbb{K}$), but which have no inflection points (over $\mathbb{K}$). The upshot of this discussion is supposed to be that not much generality is lost in restricting our attention to cubics in Weierstrass form.

The essential geometric features of Weierstrass cubics, that we have made precise in Theorem 5.10.3, allow one to endow the set $E := Z(g)^{\text{ns}}$ of non-singular points of $Z(g) \subseteq \mathbb{K}\mathbb{P}^2$ with the structure of an abelian group! For a Weierstrass cubic $g$, we let $P \mapsto \overline{P}$ denote the involution $[x : y : z] \mapsto [x : -y : z]$ of $Z(g)$. Since this involution is a projective equivalence, it takes smooth points to smooth points, so it can also be viewed as an involution of $E$. The group law $\boxplus$ on $E$ is defined as follows:

**Definition 5.10.5.** For distinct points $P, Q \in E$, let $L(P, Q) \subseteq \mathbb{K}\mathbb{P}^2$ be the unique line containing $P$ and $Q$. If $P = Q$, then $L(P, P) := T_P E$ is defined to be the tangent line to
Chapter 5 Projective Plane Curves

76

at \( P \) (here we need \( P \) to be a non-singular point). By “Bezout’s Theorem” (really: Exercise 5.13, plus Exercise 5.14 in the \( P = Q \) case), \( L(P, Q) \cap Z(g) \) consists of three points when counted with multiplicity, so we can write \( L(P, Q) \cap Z(g) = \{P, Q, R\} \) with the understanding that \( R \) might be equal to \( P, Q, \) or both. In fact the point \( R \) must be a smooth point of \( Z(g) \), otherwise, using Exercise 5.15, we would find at least four points, counted with multiplicity, in \( L(P, Q) \cap Z(g) \). We define \( P \boxplus Q := R \).

It is clear from the definition of \( \boxplus \) that \( \boxplus \) is commutative. In the interest of brevity of these notes, we shall not include here a proof that \( \boxplus \) is associative—we will now establish the other group axioms.

**Proposition 5.10.6.** Let \( g \) be a Weierstrass cubic, \( \boxplus \) the binary operation on \( E = Z(g)^{ns} \) defined in Definition 5.10.5.

1. The point \( \infty \in E \) is the identity element for \( \boxplus \). That is, \( \infty \boxplus P = P \) for all \( P \in E \).

2. For all \( P \in E \), we have \( P \boxplus P = \infty \), so \( P \) is the inverse of \( P \).

**Proof.** We saw in Proposition 5.10.2(2) that \( \infty \) is an inflection point of \( Z(g) \), hence \( L(\infty, \infty) = Z(z) \) is the line at infinity and \( L(\infty, \infty) \cap E = \{\infty, \infty, \infty\} \). We also saw in Proposition 5.10.2(4) that \( \infty = \infty \). From the definition of \( \boxplus \), it is now clear that both parts of the proposition hold when \( P = \infty \). Since \( \infty \) is the unique point of \( E \) on the line at infinity (Proposition 5.10.2(2)), we can assume in the rest of the proof that \( P = (s, t) = [s : t : 1] \) is a finite point of \( E \). Then \( L(P, \infty) = Z(x - sz) \) intersects \( E \) at the point \( \overline{P} = [s : -t : 1] \) (with multiplicity two if \( t = 0 \), in which case \( P = \overline{P} \) and \( Z(x - sz) \) is the tangent line to \( E \) at \( P \)) so we find that

\[
L(P, \infty) \cap E = \{P, \infty, \overline{P}\} \\
= L(P, \overline{P}) \cap E.
\]

The two parts of the proposition now follow immediately from the definition of \( \boxplus \). 

There is a tremendous amount to say about the group \((E, \boxplus)\), most of which is beyond the scope of these notes. We just mention the following celebrated theorem:

**Theorem 5.10.7.** *(Mordell-Weil)* If \( K = \mathbb{Q} \) (or, more generally, any number field) and \( g \in K[x, y, z] \) is a Weierstrass cubic such that \( Z(g) \) has no singular points, then the group \((E = Z(g), \boxplus)\) is finitely generated.

Some standard references for this material are [Sil] and [Tat].

5.11 Nodal and cuspital cubics

In this section we will explicitly describe the abelian group \((E, \boxplus)\) for two special Weierstrass cubics. We continue to assume that \( K \) is a field with \( 2 \in K^* \).

**Definition 5.11.1.** The Weierstrass cubic forms \( y^2z - x^2(x + z) \) and \( y^2z - x^3 \) will be called the **nodal cubic** and the **cuspital cubic**, respectively.
Note that the point \([0 : 0 : 1]\) is the unique singular point of both the nodal cubic and the cuspidal cubic, corresponding to the fact that 0 is the unique repeated root of the polynomials \(x^2(x + 1)\) and \(x^3\) (cf. Proposition 5.10.2(1)). We leave it as an exercise to prove that any Weierstrass cubic form with a singular point is projectively equivalent to either the nodal cubic or the cuspidal cubic (Exercise 5.21).

**Theorem 5.11.2.** The map \(\phi(s) := [s : 1 : s^3]\) defines an isomorphism of groups \(\phi\) from \(K\) (under addition) to \(E = Z(y^2z - x^3)^{ns}\) (under the group operation \(\boxplus\) of Definition 5.10.5).

**Proof.** Note that \([0 : 0 : 1]\) is the only singular point of \(Z(y^2z - x^3)\) and \(\infty = [0 : 1 : 0]\) is the only infinite point of \(E\); any other point of \(E\) is of the form \((x, y) = [x : y : 1]\) where \(y^2 = x^3 \neq 0\)—such a point can be written in the form \((s^{-2}, s^{-3})\) for a unique \(s \in K^*\). For such an \(s\), we have

\[
(s^{-2}, s^{-3}) = [s^{-2} : s^{-3} : 1] = [s : 1 : s^3].
\]

Since \(\phi\) takes \(s = 0\) to \(\infty\), this discussion proves that \(\phi\) is bijective and takes the identity element 0 of \((K, +)\) to the identity element \(\infty\) of \((E, \boxplus)\). It remains to check that \(\phi\) preserves addition—that is, that

\[
[s + t : 1 : (s + t)^3] = [s : 1 : s^3] \boxplus [t : 1 : t^3] \quad (5.19)
\]

for all \(s, t \in K\). This is clear if \(s\) or \(t\) is zero: If, say, \(s = 0\), then \(s = 0\) and \(\phi(s) = \infty\) are the identity elements and both sides of (5.19) are \([t : 1 : t^3]\). It is also clear if \(s = -t\), for then \(\phi(s) = \phi(t)\) and both sides of (5.19) are \(\infty\). The case where \(s = t \neq 0\) is special because we need to calculate the tangent line to \(E\) at \((s^{-2}, s^{-3})\) in order to calculate the right hand side of (5.19)—we will leave this case as Exercise 5.22 and just treat the remaining “generic” case where \(s\) and \(t\) are non-zero, \(s \neq t\), and \(s \neq -t^2\). Looking at the definition of \(\boxplus\), what needs to be shown is that the line \(L(s, t) \subseteq K^2\) containing \((s^{-2}, s^{-3})\) and \((t^{-2}, t^{-3})\) also contains the point \(((s + t)^{-2}, -(s + t)^{-3})\). As one probably knows from high school, the equation for the line containing \((x_0, y_0)\) and \((x_1, y_1)\) is

\[
y(x_0 - x_1) - x(y_0 - y_1) + y_0x_1 - x_0y_1 = 0,
\]

so the equation for \(L(s, t)\) is

\[
y(s^{-2} - t^{-2}) - x(s^{-3} - t^{-3}) + s^{-2}t^{-2}(s^{-1} - t^{-1}). \quad (5.20)
\]

Indeed, one checks easily that \((s^{-2}, s^{-3})\) and \((t^{-2}, t^{-3})\) are in the zero locus of this linear polynomial—notice that we need to know \(s \neq t\) to know that this polynomial is actually non-constant. Finally, we need to show that

\[
-(s + t)^{-3}(s^{-2} - t^{-2}) - (s + t)^{-2}(s^{-3} - t^{-3}) + s^{-2}t^{-2}(s^{-1} - t^{-1}) \quad (5.21)
\]

\[\text{When } K = \mathbb{R} \text{ or } \mathbb{C} \text{ one could argue on continuity grounds that this “generic” case implies all the other cases. In fact, one can make similar arguments over an arbitrary field } K \text{—this is one reason why I don’t feel so bad about relegating the special case } s = t \neq 0 \text{ to the exercises.} \]
is zero. Since \( s \neq -t, \ s + t \neq 0 \), so we can check that (5.21) is zero after multiplying through by

\[ (s + t)^3 = s^3 + 3s^2t + 3st^2 + t^3, \]

in which case (5.21) becomes

\[-s^{-2} + t^{-2} - (s + t)(s^{-3} - t^{-3}) + (st^{-2} + 3t^{-1} + 3s^{-1} + s^{-2}t)(s^{-1} - t^{-1}).\]

Expanding this out, one sees that it is zero.

**Theorem 5.11.3.** The map \( \phi([x : y : z]) := (y + x)/(y - x) \) defines an isomorphism of groups from \( E = \mathbb{Z}(y^2z - x^2(x + z))^{\times} \) to \( \mathbb{K}^* \) (under multiplication).

**Proof.** Exercise 5.23

### 5.12 Exercises

**Exercise 5.1.** Show that “projective equivalence” (Definition 5.3.1) defines an equivalence relation on the set of homogeneous polynomials and on the set of subsets of \( \mathbb{P}^2 \).

**Exercise 5.2.** Let \( f \in \mathbb{K}[x, y] \) be a polynomial of degree \( d \). Write \( f = \sum_{i+j \leq d} a_{i,j}x^iy^j \). Let \( \tilde{f} = \sum_{i+j \leq d} a_{i,j}x^iy^jz^{d-i-j} \) be the homogenization of \( f \) and let \( \tilde{f} := \sum_{i+j = d} a_{i,j}x^iy^j \) be the “leading order part of \( f\).”

1. Show that the intersection \( Z(\tilde{f}) \cap \mathbb{P}^1 \) of \( Z(\tilde{f}) \subseteq \mathbb{P}^2 \) with the line at infinity \( \mathbb{P}^1 \subseteq \mathbb{P}^2 \) is equal to \( Z(\tilde{f}) \), so that the intersection of the projective completion of \( Z(f) \) and the line at infinity depends only on the leading order part of \( f \). This is reasonable because the set of points “at infinity” in the projective completion of \( Z(f) \) should have to do only with the asymptotic behaviour of \( f \).

2. Calculate \( Z(\tilde{f}) \cap \mathbb{R}^1 \) when

\[ f = x^3 - y^3 + 5x^2 - 7xy + 5x + 13 \in \mathbb{R}[x, y] \]

What is the cardinality of \( Z(\tilde{f}) \cap \mathbb{R}^1 \)? What would change if we replaced the real numbers with the complex numbers everywhere and asked for the cardinality of \( Z(\tilde{f}) \cap \mathbb{C}^1 \)?

**Exercise 5.3.** State and prove the analogue of Exercise 3.2 for a homogeneous polynomial \( f \in \mathbb{K}[x, y, z] \).

**Exercise 5.4.** Let \( f \in \mathbb{K}[x, y, z] \) be a homogeneous polynomial, \( A \in \text{GL}_3(\mathbb{K}) \) an invertible matrix. Prove that a point \( P \in Z(f) \subseteq \mathbb{P}^2 \) is a singular point of \( Z(f) \) iff \( A^{-1}(P) \) is a singular point of \( Z(f \cdot A) \). (Cf. Exercise 3.3.)

**Exercise 5.5.** Prove Lemma 5.4.1.
Exercise 5.6. Let \( \mathbb{K} \) be a field of characteristic 2. Show that the homogeneous degree two polynomial \( xy \in \mathbb{K}[x, y] \) is not projectively equivalent to one of the form \( Ax^2 + By^2 \).

Exercise 5.7. Let \( f(x) \in \mathbb{K}[x] \) be a polynomial in one variable over a field \( \mathbb{K} \). The multiplicity of a root \( \alpha \) of \( f \) is defined in §5.8.

1. If \( \alpha \) is a root of \( f \) with multiplicity \( m \), show that
\[
f(\alpha) = f'(\alpha) = \cdots = f^{(m-1)}(\alpha) = 0,
\]
but \( f^{(m)}(\alpha) \neq 0 \). (This gives an alternative definition of \( m \).)

2. Show that if \( f \) has at least \( d - 1 \) roots in \( \mathbb{K} \), counted with multiplicity (that is, at least \( d - 1 \) of the \( \alpha_i \) are in \( \mathbb{K} \) for a factorization of \( f \) in \( \mathbb{K}[x] \) as in (5.10)), then in fact \( f \) has \( d \) roots in \( \mathbb{K} \), counted with multiplicity (that is, all the \( \alpha_i \) are in \( \mathbb{K} \), so the factorization of \( f \) above can actually be done in \( \mathbb{K}[x] \)). The point here is that, when we write \( f = C(x^d + a_{d-1}x^{d-1} + \cdots + a_0) \), the coefficients \( a_{d-1}, \ldots, a_0 \) are in \( \mathbb{K} \), but they are also expressible somehow (how?) in terms of the \( \alpha_i \).

3. Clearly \( m(\lambda f, \alpha) = m(f, \alpha) \) for any \( \lambda \in \mathbb{K}^* \). Show also that \( m(f(\lambda x), \lambda^{-1}\alpha) = m(f, \alpha) \) for \( \lambda \in \mathbb{K}^* \) and \( m(f(x + a), \alpha - a) = m(f, \alpha) \) for any \( a \in \mathbb{K} \). That is, “multiplicity is invariant under \( \text{Aff}(\mathbb{K}) \).”

Exercise 5.8. Suppose \( f \in \mathbb{K}[u, v] \) is a homogeneous polynomial of degree \( d > 0 \) in variables \( u, v \) and \( P = [s : t] \in Z(f) \subseteq \mathbb{K}P^1 \). If \( P \in U_0 \subseteq \mathbb{K}P^1 \) (resp. \( P \in U_1 \)), we define the multiplicity \( m = m(f, P) \) of \( P \) (as a point of \( Z(f) \)) to be the multiplicity of \( t/s \) (resp. \( s/t \)) as a root of \( g(x) := f(1, x) \in \mathbb{K}[x] \) (resp. \( h(x) := f(x, 1) \in \mathbb{K}[x] \)). Show that if \( P \in U_0 \cap U_1 \), then both ways of defining \( m \) give the same result, so this actually makes sense!

Exercise 5.9. Let \( f \) and \( P \) be as in the previous exercise and let
\[
A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathbb{K})
\]
be an invertible \( 2 \times 2 \) matrix. Let \( \overline{A} : \mathbb{K}P^1 \to \mathbb{K}P^1 \) be the associated projective transformation defined by \( \overline{A}([x]) := [Ax] \). Show that \( m(f, P) = m(f \cdot A, \overline{A}^{-1}(P)) \). Recall that
\[
(f \cdot A)(x, y) = f(au + bv, cu + dv).
\]
In other words, multiplicity is invariant under projective transformations of \( \mathbb{K}P^1 \).

Exercise 5.10. Now let \( g \in \mathbb{K}[x, y, z] \) be a homogeneous polynomial of degree \( d > 0 \), let \( L \subseteq \mathbb{K}P^2 \) be a line, and let \( P \) be a point of \( L \cap Z(g) \). We want to define the multiplicity \( m = m(P, g, L) \) of the intersection \( L \cap Z(g) \) at \( P \) when it makes sense to do so. As we saw on a previous homework, \( L = \mathbb{P}L \) for a unique two dimensional linear subspace \( \tilde{L} \subseteq \mathbb{K}^3 \). Choose an ordered basis \( (q, r) \) for \( \tilde{L} \) and set
\[
f(u, v) := g(uq + vr).
\]
More precisely, if \( q = (q_1, q_2, q_3) \) and \( r = (r_1, r_2, r_3) \), then
\[
 f(u, v) := g(uq_1 + vr_1, uq_2 + vr_2, uq_3 + vr_3).
\]

We also have a bijection
\[
 \mathbb{K}P^1 \to L
\]
\[
 [u, v] \mapsto [uq + vr],
\]
so we can write \( P = [sq + tr] \) for a unique \([s, t] \in \mathbb{K}P^1 \). The polynomial \( f(u, v) \in \mathbb{K}[u, v] \) is clearly “of degree \( d \),” except that it might be zero. (Note that if \( f \) is zero, then \( L \subseteq Z(g) \).)

Assuming that \( L \) is not contained in \( Z(g) \), so \( f \) is not zero, we define \( m(P, g, L) \) to be the multiplicity of \([s, t] \) as a point of \( Z(f) \) in the sense of Exercise 5.8. Show that \( m(P, g, L) \) is independent of the choice of ordered basis \((b, c)\) for \( V \).

**Exercise 5.11.** Prove Lemma 5.2.2.

**Exercise 5.12.** Let \( f \in \mathbb{K}[x, y, z] \) be a homogeneous polynomial of degree \( d \), so that \( f(x, y, 1) \in \mathbb{K}[x, y] \) is a polynomial of degree \( d \), though not necessarily homogeneous. Show that a point \( P = (x_0, y_0) \) of the affine algebraic curve \( Z(f(x, y, 1)) \subseteq \mathbb{K}^2 \) is a smooth point in the sense of Definition 3.2.1 if the corresponding point \([x_0 : y_0 : 1] \) (also abusively denoted \( P \)) is a smooth point of the projective curve \( Z(f) \subseteq \mathbb{K}P^2 \) in the sense of Definition 5.2.1. Furthermore, if these equivalent smoothness notions hold, show that the tangent line to \( Z(f) \subseteq \mathbb{K}P^2 \) at \( P \) defined in Definition 5.2.3 is the projective completion of the tangent line to \( Z(f(x, y, 1)) \subseteq \mathbb{K}^2 \) at \( P \) defined in Definition 3.2.2.

**Exercise 5.13.** Let \( g \in \mathbb{K}[x, y, z] \) be a (non-zero) homogeneous polynomial of degree 3, \( L \subseteq \mathbb{K}P^2 \) a line. Assume that \( L \) is not contained in \( Z(g) \). Show that if \( L \cap Z(g) \) contains at least two points (counted with multiplicity), then it contains precisely three points (counted with multiplicity).

**Exercise 5.14.** Let \( f \in \mathbb{K}[x, y, z] \) be a form, \( P \) a smooth point of \( Z(f) \). Show that the intersection multiplicity \( m(T_P Z(f), f, P) \) of the tangent line \( T_P Z(f) \) and \( Z(f) \) at \( P \) is at least two.

**Exercise 5.15.** Let \( f \in \mathbb{K}[x, y, z] \) be a form, \( P \) a singular point of \( Z(f) \), \( L \) a line in \( \mathbb{K}P^2 \) containing \( P \). Show that the intersection multiplicity \( m(L, f, P) \) of \( L \) and \( Z(f) \) at \( P \) is at least two.
Exercise 5.16. Prove Pappus’ Theorem (Theorem 5.9.2).

Exercise 5.17. Prove Desargues’ Theorem (Theorem 5.9.3).

Exercise 5.18. Suppose $A, B, C, A', B', C' \in \mathbb{R}^2 \subseteq \mathbb{RP}^2$ are six distinct points in general position, except that $A, B, \text{ and } C$ are contained in a line $L$. Notice that condition (2) in Desargues’ Theorem (Theorem 5.9.3) holds trivially since $X, Y, Z \in L$. Draw such a configuration where (1) does not hold.

Exercise 5.19. Consider the cubic polynomial $f = f(x) = x(x^2 + 1) \in \mathbb{R}[x]$ and the associated Weierstrass cubic $g = g(x, y) = y^2 - f(x)$. Let $E := Z(\tilde{g}) \subseteq \mathbb{RP}^2$, $\infty := [0 : 1 : 0] \in E$. Recall the binary operation $\boxplus$ from Definition 5.10.5 making $(E, \boxplus)$ an abelian group.

a) Find all points $P$ of $E$ for which $P \boxplus P = \infty$.

b) Calculate $\infty \boxplus [1 : \sqrt{2} : 1]$.

c) Calculate $[0 : 0 : 1] \boxplus [1 : \sqrt{2} : 1]$.


Exercise 5.20. Let $g = y^2z - \tilde{f}(x, z)$ be a Weierstrass cubic form. Show that a point $P \in E = Z(\tilde{g})$ has order two in $(E, \boxplus)$ (meaning $P \neq \infty$, but $P \boxplus P = \infty$) iff $P = (\alpha, 0)$ for a root $\alpha$ of $f$.

Exercise 5.21. Suppose $g = y^2x - \tilde{f}(x, z)$ is a Weierstrass cubic form (Definition 5.10.1) over a field $\mathbb{K}$ such that $Z(g) \subseteq \mathbb{KP}^2$ has a singular point (over $\mathbb{K}$). Prove that $g$ is projectively equivalent to either the nodal cubic or the cuspidal cubic (Definition 5.11.1). Prove that the nodal cubic and the cuspidal cubic are not projectively equivalent.

Exercise 5.22. For the cuspidal cubic $g = y^2z - x^3$, calculate the tangent line $L(s, s)$ to $Z(g)$ at the point $(s^{-2}, s^{-3})$ ($s \in \mathbb{K}^*$) and use your formula for $L(s, s)$ to check equation (5.19) in the proof of Theorem 5.11.2 in the case $s = t$.

Exercise 5.23. Prove Theorem 5.11.3.
In this appendix I have collected together some basic facts, mostly of algebraic nature, that are used in the text. My hope is that these notes should be readable without this appendix, but I have included this anyway for the reader who wants to have a better idea of the “general context” of some constructions in the main text.

6.1 Groups and group actions

Definition 6.1.1. A group \((G, \circledast)\) is a set \(G\) equipped with a binary operation \((g, h) \mapsto g \circledast h\) satisfying the following properties:

1. \(\circledast\) is associative: \((g \circledast h) \circledast k = g \circledast (h \circledast k)\) for all \(g, h, k \in G\).

2. \(\circledast\) has an identity: There is an element \(e \in G\) (necessarily unique; denoted \(e_G\) if there is any chance of confusion) such that \(e \circledast g = g \circledast e = g\) for all \(g \in G\).

3. \(\circledast\) has inverses: For every \(g \in G\) there is an \(h \in G\) such that \(g \circledast h = h \circledast g = e\).

A group \((G, \circledast)\) is called abelian (or commutative) iff \(g \circledast h = h \circledast g\) for all \(g, h \in G\). If \((G, \boxplus)\) and \((G', \boxplus')\) are groups, a group homomorphism (or simply, a map of groups) from \((G, \boxplus)\) to \((G', \boxplus')\) is a function \(f : G \rightarrow G'\) such that \(f(e_G) = e_G'\) and \(f(g \boxplus h) = f(g) \boxplus' f(h)\) for all \(g, h \in G\). An isomorphism of groups is a bijective group homomorphism.

It is customary to use “\(G\)” to refer both to a group \((G, \boxplus)\) and the “underlying set” \(G\). Usually, when discussing an arbitrary group \(G\), one usually denotes the binary operation \(\boxplus\) simply by juxtaposition, writing “\(gh\)” instead of “\(g \boxplus h\).” Similarly, it is customary to denote the identity element \(e\) of \(G\) by \(1\), or perhaps \(1_G\) if there is any chance of confusion. For \(g \in G\), the element \(h \in G\) satisfying \(gh = hg = 1\) is easily seen to be unique, and is usually denoted \(g^{-1}\) and called the inverse of \(g\).

For an abelian group, one usually denotes the binary operation by \((g, h) \mapsto g + h\), the identity element by \(0\), and the inverse of \(g\) by \(-g\).

Definition 6.1.2. For a group \(G\), a non-empty subset \(H \subseteq G\) is called a subgroup iff it satisfies:

82
1. \( hh' \in H \) for all \( h, h' \in H \)
2. \( h^{-1} \in H \) for all \( h \in H \).

These conditions are usually summarized by saying that \( H \) is \textit{closed under addition and inverses}.

A subgroup of a group is a group in its own right, with the binary operation inherited from \( G \). If \( f : G \to G' \) is a map of groups and \( H \) is a subgroup of \( G \), then one checks easily that the image \( f(H) \) of \( H \) under \( f \) is a subgroup of \( G' \). In particular, the image \( f(G) \) of \( G \) is a subgroup of \( H \).

Example 6.1.3. For a group \( G \) and an element \( g \in G \), the function \( f_g : G \to G \) given by \( f_g(h) := ghg^{-1} \) is an isomorphism of groups called \textit{conjugation by} \( g \). In particular, if \( H \) is a subgroup of \( G \), then the image

\[
   f_g(H) = \{ ghg^{-1} : h \in H \}
\]

(also denoted \( gHg^{-1} \)) of \( H \) under \( f_g \) is also a subgroup of \( G \).

Definition 6.1.4. Two subgroups \( H \) and \( H' \) of a group \( G \) are called \textit{conjugate} iff there is some \( g \in G \) such that \( H' = gHg^{-1} \). A subgroup \( N \) of \( G \) is called \textit{normal} iff \( N \) is the only subgroup of \( G \) conjugate to \( N \)—equivalently \( N = gNg^{-1} \) for every \( g \in G \).

Definition 6.1.5. If \( G \) is a group and \( S \) is a set, then an \textit{action} (more precisely: \textit{left action}) of \( G \) on \( S \) is a function

\[
   G \times S \to S \\
   (g, s) \mapsto gs
\]

satisfying:

1. \( 1s = s \) for all \( s \in S \). Here \( 1 \in G \) is the identity element of \( G \).
2. \( g(hs) = (gh)s \) for all \( g, h \in G, s \in S \).

Let \( S \) be a set equipped with an action of a group \( G \) (a \textit{“\( G \)-set”}). For \( s \in S \), the subset

\[
   \text{Stab} s := \{ g \in G : gs = s \}
\]

of \( G \) is called the \textit{stabilizer} of \( s \) (in \( G \)) and the subset

\[
   Gs := \{ gs : g \in G \}
\]

of \( S \) is called the \textit{orbit} (or \textit{G-orbit} if there is a chance of confusion) of \( s \).

Proposition 6.1.6. For any \( s \in S \), the stabilizer \( \text{Stab} s \) is a subgroup of \( G \). If \( s, t \in G \) have the same orbit, then \( \text{Stab} s \) and \( \text{Stab} t \) are conjugate subgroups of \( G \).

\textit{Proof.} Exercise 6.2
6.2 Fields and division rings

**Definition 6.2.1.** A *right near-field* is a set $\mathbb{K}$ equipped with two binary operations: “addition,” denoted $(x, y) \mapsto x + y$ and “multiplication,” denoted $(x, y) \mapsto xy$, satisfying:

1. $(\mathbb{K}, +)$ is an abelian group, called the *additive group* of $\mathbb{K}$. Let $0 \in \mathbb{K}$ denote its identity element.

2. $\mathbb{K}^* := \mathbb{K} \setminus \{0\}$ is a group under multiplication, called the *multiplicative group* of $\mathbb{K}$. The identity element for this group is denoted 1, or perhaps $1_\mathbb{K}$ if there is any chance of confusion.

3. Multiplication on the right distributes over addition:

$$ (x + y)z = xz + yz \quad (6.1) $$

for all $x, y, z \in \mathbb{K}$.

A right near field is called a *division ring* iff it satisfies:

3. Multiplication on the left distributes over addition:

$$ x(y + z) = xy + xz \quad (6.2) $$

for all $x, y, z \in \mathbb{K}$.

A *field* is a division ring $\mathbb{K}$ whose multiplicative group $\mathbb{K}^*$ is abelian. An *isomorphism* $f : \mathbb{K} \to \mathbb{K}'$ of near-fields, division rings, or fields is a bijective function with $f(0) = 0$, $f(1) = 1$, and satisfying $f(x + y) = f(x) + f(y)$ and $f(xy) = f(x)f(y)$ for all $x, y \in \mathbb{K}$.

For example, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are fields with the usual notions of addition and multiplication. For each prime number $p$, the set

$$ \mathbb{F}_p := \mathbb{Z}/p\mathbb{Z} $$

of integers mod $p$ is a field, with addition and multiplication given by the usual addition and multiplication of integers, modulo $p$.

If $\mathbb{K}$ is a field, a subset $\mathbb{K}' \subseteq \mathbb{K}$ of $\mathbb{K}$ is called a *subfield* of $\mathbb{K}$ iff $\mathbb{K}'$ is closed under addition and multiplication in $\mathbb{K}$ and these operations make $\mathbb{K}'$ into a field. It is clear that any intersection of subfields is again a subfield, hence any subset $S$ of a field $\mathbb{K}$ is contained in a smallest subfield of $\mathbb{K}$ (construct it by intersecting all subfields of $\mathbb{K}$ containing $S$). If $\mathbb{K}'$ is a subfield of $\mathbb{K}$, then $\mathbb{K}$, with its additive abelian group structure, becomes a vector space over $\mathbb{K}'$ by defining scalar multiplication as multiplication in $\mathbb{K}$.

**Definition 6.2.2.** A *number field* is a subfield of $\mathbb{C}$ (any such subfield contains $\mathbb{Q}$) which is finite dimensional as a $\mathbb{Q}$ vector space.
Example 6.2.3. The smallest subfield $\mathbb{Q}(\sqrt{2})$ of $\mathbb{C}$ containing $\sqrt{2}$ is the set of all real numbers $r$ that can be written in the form $r = a + b\sqrt{2}$ for some $a, b \in \mathbb{Q}$. Since $\sqrt{2} \notin \mathbb{Q}$, we have $a + b\sqrt{2} = a' + b'\sqrt{2}$ for $a, a', b, b' \in \mathbb{Q}$ iff $a = a'$ and $b = b'$. Hence $\sqrt{2}$ is a basis for $\mathbb{Q}(\sqrt{2})$ as a $\mathbb{Q}$ vector space, so $\mathbb{Q}(\sqrt{2})$ is a number field.

Definition 6.2.4. The characteristic of a field $K$ is defined to be the smallest positive integer $p$ such that $p = 1 + \cdots + 1$ ($p$ times) is equal to zero in $K$; if there is no such positive integer, the characteristic of $K$ is defined to be zero.

From the fact that $K^* = K \setminus \{0\}$ is a group, it follows that the characteristic of $K$ is prime (when non-zero). Any field of characteristic zero contains a subfield isomorphic to $\mathbb{Q}$. Any field of characteristic $p$ contains a subfield isomorphic to $\mathbb{F}_p$. If $K$ is a field of characteristic $p$ and $q = p^n$ is a prime power, then $f(x) := x^q$ defines an automorphism $f : K \to K$ called the Frobenius automorphism of order $q$.

Here is a summary of some basic notions and results from the theory of fields. For all of the results on finite fields presented here, I recommend [Her, §7.1] as a reference.

Definition 6.2.5. A field $K$ is called algebraically closed iff any polynomial $f(x) \in K[x]$ of positive degree can be factored as

$$f(x) = C \prod_{i=1}^{d} (x - \alpha_i)$$

for some $C, \alpha_1, \ldots, \alpha_d \in K$.

Theorem 6.2.6. Every field is a subfield of an algebraically closed field.

Proof. This is perhaps the most fundamental result in the theory of fields. This is proved in almost every abstract algebra textbook, though I can’t find it explicitly stated in [Her] as Herstein likes to stick to finite field extensions at all times (it does follow the existence and “uniqueness” of splitting fields for polynomials—Theorems 5.H and 5.J in [Her]). One standard reference is [DF, Proposition 30, §13.4].

Theorem 6.2.7. The number of element $q := \#K$ in any finite field $K$ is a prime power: $q = p^n$. For each prime power $q = p^n$ there is a unique (up to isomorphism) finite field $\mathbb{F}_q$ with $q$ elements. The $q$ elements of $\mathbb{F}_q$ are the roots of the polynomial $x^q - x \in \mathbb{F}_p[x]$, which has coefficients in $\mathbb{F}_p \subseteq \mathbb{F}_q$. If $r = p^m$ for some $m \geq n$, then $\mathbb{F}_q$ is a subfield of $\mathbb{F}_r$, equal to the fixed locus of the Frobenius automorphism of $\mathbb{F}_r$ of order $q$.

Proof. This is a summary of the first couple of results in [Her, §7.1].

Theorem 6.2.8. (Wedderburn) Any finite division ring is a field.

Proof. See, for example, [Her, Theorem 7.C] (where there are two proofs) or [DF, Exercise 13, §13.6]. Wedderburn’s original proof (1905) had a slight gap; the first complete proof is usually attributed to Leonard Dickson. The most common “modern proof” that one encounters is due to Ernst Witt.

Theorem 6.2.9. The multiplicative group $K^*$ of any finite field $K$ is cyclic.
Proof. See, for example, [Her, Theorem 7.B].

Example 6.2.10. The Dickson near-field $J_9$ is a near-field defined as follows. As an additive group, $(J_9, +)$ is “the” field $\mathbb{F}_9$ of order 9: $(J_9, +) := (\mathbb{F}_9, +)$. The multiplication $\ast$ for $J_9$ is defined in terms of the multiplication on $\mathbb{F}_9$ (which we denote by juxtaposition, as usual) by the rule

$$a \ast b := \begin{cases} \newline ab, & b \in S \\ \newline a^3b, & b \notin S, \end{cases}$$

where $S := \{a^2 : a \in \mathbb{F}_9\}$ is the set of squares in $\mathbb{F}_9$.

For more on near-fields and related algebraic structures, see the survey article [Wei] by Charles Weibel.

6.3 The Implicit Function Theorem

Here is a version of the Implicit Function theorem which is not the most general statement possible, but which is sufficient for our purposes:

**Theorem 6.3.1. (Implicit Function Theorem)** Let $U$ be an open subset of $\mathbb{R}^{n+1}$, $f : U \rightarrow \mathbb{R}$ a continuously differentiable function (for example, a polynomial function), $P = (a_1, \ldots, a_n, b)$ a point of $U$. Assume that the partial derivative of $f$ with respect to $x_{n+1}$ is non-zero at $P$. Then there are open subsets $V \subseteq \mathbb{R}^n$ and $W \subseteq \mathbb{R}$ containing $(a_1, \ldots, a_n)$ and $f(P)$, respectively, and a continuously differentiable function $g : V \rightarrow W$ such that the graph of $g$,

$$\Gamma_g = \{(x_1, \ldots, x_n, g(x_1, \ldots, x_n)) : (x_1, \ldots, x_n) \in V\},$$

is equal to the level set

$$(f|V)^{-1}(f(P)) = \{(x_1, \ldots, x_{n+1}) \in V \times W : f(x_1, \ldots, x_{n+1}) = f(P)\}$$

of $f$ on $V$. Furthermore, if $f$ is $k$ times continuously differentiable on $V \times W$, then so is $g$.

The role played by the “last” coordinate in the Implicit Function Theorem is nothing special—this coordinate is singled out just to make the statement of the theorem as clean as possible. Usually one applies the theorem to a differentiable function $f : U \rightarrow \mathbb{R}$ on an open subset $U \subseteq \mathbb{R}^{n+1}$ for which it is known that $f_i(P) \neq 0$ for some partial derivative $f_i$ of $f$. One concludes that there is an open neighborhood $V$ of $P$ in $\mathbb{R}^{n+1}$ such that $V \cap f^{-1}(f(P))$ is mapped bijectively to an open subset $W \subseteq \mathbb{R}^n$ via the projection

$$\pi_i : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n$$

$$\pi_i(x_1, \ldots, x_{n+1}) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1})$$

and, furthermore, the inverse of $\pi_i : V \cap f^{-1}(f(P)) \rightarrow W$ is given by a function $g : W \rightarrow V \cap f^{-1}(f(P)) \subseteq \mathbb{R}^{n+1}$ which is “at least as differentiable as $f$.”

There is also an analytic version of the implicit function theorem:
Theorem 6.3.2. Let $U$ be an open subset of $\mathbb{C}^{n+1}$, $f : U \to \mathbb{C}$ an analytic function on $U$ (a function equal to a convergent series Taylor series on a neighborhood of each point of $U$—for example, a function given by a polynomial with complex coefficients, or a ratio of two such polynomials whose denominator is non-zero on $U$), $P = (a_1, \ldots, a_n, b)$ a point of $U$. Assume that the partial derivative $f_{z_{n+1}}(P)$ is non-zero. Then there are open subsets $V \subseteq \mathbb{C}^{n}$ and $W \subseteq \mathbb{C}$ containing $(a_1, \ldots, a_n)$ and $f(P)$, respectively, and an analytic function $g : V \to W$ such that the graph of $g$,

$$\Gamma_g = \{ (z_1, \ldots, z_n, g(z_1, \ldots, z_n)) : (z_1, \ldots, z_n) \in V \},$$

is equal to the level set 

$$(f|V)^{-1}(f(P)) = \{ (z_1, \ldots, z_{n+1}) \in V \times W : f(z_1, \ldots, z_{n+1}) = f(P) \}$$

of $f$ on $V$.

It is worth emphasizing that in both of these “implicit function theorems,” the inverse function $g$ will rarely be a polynomial function (or even a rational function) even when the “input” function $f$ is a polynomial.

Example 6.3.3. The function $f(x, y) := y^2 - x^3$ is a polynomial (hence infinitely differentiable) function $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f_x = 3x^2$ is non-zero at $P = (1, 1) \in f^{-1}(0)$. Consider the open neighborhood $V := \mathbb{R}_{>0}^2$ of $P$ in $\mathbb{R}^2$ and the open subset $W := \mathbb{R}_{>0}$ of $\mathbb{R}$. Then the projection $\pi(x, y) := x$ defines a bijection $\pi : V \cap f^{-1}(0) \to W$ with inverse $g(x) := (x, \sqrt[3]{x^3})$. Although $g : W \to \mathbb{R}^2$ is a perfectly nice infinitely differentiable function, it is not given by a ratio of polynomials on any neighborhood of $\pi(P) = 1$.

6.4 Exercises

Exercise 6.1. Let $G$ be a group. Show that “being conjugate” (Definition 6.1.4) is an equivalence relation on the set of subgroups of $G$.


Exercise 6.3. Check that the Dickson near field $J_9$ defined in Example 6.2.10 is actually a near-field (Definition 6.2.1). Show that $J_9$ does not satisfy the left distributive law (6.2), so that it is not a division ring.

Exercise 6.4. Show, by writing down an explicit isomorphism, that the multiplicative group $J_9^*$ of the Dickson near field $J_9$ (Example 6.2.10) this group is isomorphic to the quaternionic group

$$Q := \{ \pm 1, \pm i, \pm j, \pm k \}.$$

(Note that, up to isomorphism, there are exactly two non-abelian groups of order 8: $Q$ and the dihedral group $D_4$—you could probably show that $J_9^*$ is isomorphic to $Q$ very easily by appealing to this classification.) Also note that the fact that $J_9^*$ is not abelian, together with Wedderburn’s Theorem (Theorem 6.2.8), gives another proof that $J_9$ cannot be a division ring.
Bibliography


