Introduction to Algebraic Geometry

Sudhir R. Ghorpade

Department of Mathematics
Indian Institute of Technology Bombay
Powai, Mumbai 400 076, India

E-Mail: `srg@math.iitb.ac.in`
URL: `http://www.math.iitb.ac.in/~srg/`

Expanded version of the notes of a lecture for beginners
at the

Science Special Lecture Series
of the
Karnataka Academy of Science and Technology
Department of Mathematics
University of Mysore

March 5, 2007

Short Term Training Programme
on
Advances in Control Theory
Department of Electrical Engineering
VJTI, Mumbai

October 4, 2007

Contents

1 Preamble

It is a hazardous task to try to give an introduction to any significant area of mathematics within the short span of a single lecture. This is particularly so for a subject like algebraic geometry, which has developed over a period of at least 200 years, and is closely related to almost every branch of mathematics, to wit, algebra, analysis, geometry, topology and number theory. In the preface to his book [20], Kunz wrote in 1985 that at the present state of our knowledge, one could give a 200 semester course on algebraic geometry (and commutative algebra) without ever repeating oneself. Talking about elliptic curves, which is one of the topics in algebraic geometry, Lang once wrote that "it is possible to write endlessly" (and followed it up by clarifying that "this is not a threat!").

It should be clear, therefore, that any brief introduction to algebraic geometry has to be selective and can at best hope to provide some glimpses of the subject. This is what we have set out to do. In fact, we will focus mainly on two basic results in algebraic geometry, known as Bezout's Theorem and Hilbert's Nullstellensatz, each of which can be viewed as a generalization of the Fundamental Theorem of Algebra. We will review some background material and try to motivate the two results and state them precisely. For proofs and further details, we will give a number of references. In fact, there are suggestions for further reading at the end of each section and also in the last section. We hope the reader will feel interested enough to look these up and make an attempt to learn more.

2 Theory of Equations

The theory of equations is concerned with solving polynomial equations. In high-school, Algebra or *Beejganit* is almost synonymous with the art of formulating, manipulating and solving polynomial equations. We learn some basic techniques and work out a variety of examples, usually restricted to polynomials in one variable of a reasonably small degree. In college, Algebra appears to be mainly the study of abstract algebraic structures such as groups, rings, fields, and we learn a number of basic notions and results concerning these objects. Slowly, the ideas of theory of equations make a reappearance in the guise of notions such as euclidean domains, principal ideal domains and unique factorization domains. Also, one revisits the familiar formula for the roots of a quadratic equation in terms of its coefficients, and investigates if such a thing is possible for equations of higher degree. This culminates in a remarkable part of algebra, known as Galois Theory.

For our purpose, it should suffice at the moment to stick to high-school algebra, and review some basic aspects of the theory of equations. However, we work in a little more generality by considering not only equations with rational or real coefficients, but equations with coefficients in a field. A *field* is basically a set in which we can add, subtract, multiply and divide. Of course division is only permissible by nonzero elements. Some examples of fields are: the set $\mathbb{Q}$ of all rational numbers, the set $\mathbb{R}$ of all real numbers, the set $\mathbb{C}$ of all complex numbers, and the set $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ of residue classes of integers modulo a prime number p. If you are uncomfortable with the notion of a field, please think of $\mathbb{C}$ whenever a field is being talked about.

Let K be a field and let $K[X]$ denote the set of all polynomials in the variable X with coefficients in K. Elements of $K[X]$ look like

$$f(X) = a_dX^d + a_{d-1}X^{d-1} + \cdots + a_1X + a_0 \quad \text{where } d \geq 0 \text{ and } a_0, a_1, \ldots, a_d \in K;$$

if $a_d \neq 0$, we say that the *degree* of $f(X)$ is d and write $\deg f(X) = d$. In this case a_d is called the *leading coefficient* of $f(X)$. Polynomials whose leading coefficient is 1 are said to be *monic*. If $a_0 = a_1 = \cdots = a_d = 0$, that is, if all the coefficients are zero, then $f(X)$ is the *zero polynomial* and we write $f(X) = 0$. Note that we have natural and familiar notions of addition, subtraction and multiplication of polynomials, which make $K[X]$ a ring. Division for polynomials is not possible, in general. The following result, which formalizes the process of long division in high-school, tells us the best that we can do, in general.

Proposition 2.1 (Division Algorithm). *Given any $f(X), g(X) \in K[X]$ with $g(X) \neq 0$, there exist unique polynomials $q(X), r(X) \in K[X]$ such that*

$$f(X) = q(X)g(X) + r(X) \quad \textit{and either } r(X) = 0 \textit{ or } \deg r(X) < \deg g(X).$$

Proving the Division Algorithm is not difficult—it suffices to assume that $f(X) \neq 0$ and induct on $\deg f(X)$ to show the existence of $q(X)$ and $r(X)$ satisfying the desired properties. The uniqueness follows by noting if one had another representation $f(X) = \tilde{q}(X)g(X) + \tilde{r}(X)$, then the difference $r(X) - \tilde{r}(X)$ is divisible by $g(X)$ and has degree $< \deg g(X)$. Thus $r(X) - \tilde{r}(X) = 0$, and as a consequence, $q(X) - \tilde{q}(X) = 0$. With this in view, one calls $q(X)$ the *quotient* and $r(X)$ the *remainder* upon division of $f(X)$ by $g(X)$. When $r(X) = 0$, we say that $g(X)$ divides $f(X)$.

Given a polynomial, one can substitute things in it. More precisely, given any $f(X) = a_dX^d + \cdots + a_1X + a_0 \in K[X]$ and any $\alpha \in K$, the element

$f(\alpha) := a_d\alpha^d + \cdots + a_1\alpha + a_0$ of K is the result of substituting α in $f(X)$. Substitution respects addition and multiplication of polynomials, that is $(f+g)(\alpha) = f(\alpha)+g(\alpha)$ and $fg(\alpha) = f(\alpha)g(\alpha)$ for any $f(X), g(X) \in K[X]$. In other words the map $\phi_\alpha : K[X] \to K$ defined by $\phi_\alpha(f(X)) = f(\alpha)$ is a ring homomorphism. We say that α is a *root* or a *zero* of $f(X)$ if $f(\alpha) = 0$. A fancy way to say this is that $f(X)$ is in the kernel of the substitution homomorphism ϕ_α.

The Division Algorithm has following useful consequences.

1. **[Remainder Theorem]** *Given any $f(X) \in K[X]$ and $\alpha \in K$, the remainder upon the division of $f(X)$ by $(X - \alpha)$ is the effect of substituting α in $f(X)$. In other words, $f(X) = q(X)(X - \alpha) + f(\alpha)$ for some $q(X) \in K[X]$.*

2. **[Factor Theorem]** *Given any $f(X) \in K[X]$, an element α of K is a root of $f(X)$ if and only if $(X - \alpha)$ divides $f(X)$.*

3. *A nonzero polynomial of degree n in $K[X]$ has at most n roots in K.*

It may be noted that the Remainder Theorem is an easy consequence of the Division Algorithm. The Factor Theorem follows readily from the Remainder Theorem. Assertion # 3 follows from the Factor Theorem by induction on $\deg f(X)$. It may also be noted that if L is any field containing K as a subfield, then a polynomial in $K[X]$ can be regarded as a polynomial in $L[X]$ and, in particular, we can substitute elements of L in it. With this in view, we also see that a nonzero polynomial of degree n in $K[X]$ has at most n roots in L. For example, a polynomial of degree n with coefficients in $\mathbb{Q}$ not only has at most n roots in $\mathbb{Q}$, but also, at most n roots in $\mathbb{R}$, and at most n roots in $\mathbb{C}$. On the other hand, a polynomial in $K[X]$ of degree n may not have n roots in K. This can always happen for a trivial reason. For example, the polynomial X^n has only one root, namely $X = 0$. We better discount this possibility by learning to *count properly*! So let us define the *multiplicity* of $\alpha \in K$ as a root of a nonzero polynomial $f(X) \in K[X]$ to be the unique nonnegative integer m such that $f(X) = (X - \alpha)^m g(X)$ for some $g(X) \in K[X]$ with $g(\alpha) \neq 0$. Now assertion # 3 can be stated in a more precise form as follows.

3′ *A nonzero polynomial of degree n in $K[X]$ has at most n roots in K, counting multiplicities. More precisely, if $f(X) \in K[X]$ is a nonzero polynomial of degree n, then it has finitely many roots $\alpha_1, \dots, \alpha_r$ in K and if e_i is the multiplicity of α_i as a root of $f(X)$, for $i = 1, \dots, r$, then $e_1 + \cdots + e_r \leq n$.*

It is natural to ask when a polynomial of degree n has exactly n roots, counted properly. Again, this does not happen, in general. For example, the polynomial $X^2 - 2$ of degree 2 in $\mathbb{Q}[X]$ has no roots in $\mathbb{Q}$. It has, however, 2 roots in the bigger field $\mathbb{R}$. The polynomial $X^2 + 1$ in $\mathbb{R}[X]$ has no roots in $\mathbb{R}$, but has two roots in $\mathbb{C}$ that are usually denoted by i and $-i$. In fact, $\mathbb{C}$ is obtained by adjoining the roots of $X^2 + 1$ to the field $\mathbb{R}$. In other words, to pass from $\mathbb{R}$ to $\mathbb{C}$, we simply arrange that the equation $X^2+1 = 0$ can be solved. But now, a miracle happens. Namely, every equation can be solved! More precisely, we have the following.

Proposition 2.2 (Fundamental Theorem of Algebra). *If $f(X)$ is a nonzero polynomial of degree n in $\mathbb{C}[X]$, then $f(X)$ has exactly n roots in $\mathbb{C}$, counting multiplicities. In other words, there are distinct complex numbers $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$ and positive integers $e_1, \ldots, e_r$ such that*

$$f(X) = a_n(X - \alpha_1)^{e_1} \cdots (X - \alpha_r)^{e_r} \quad \textit{with} \quad e_1 + \cdots + e_r = n,$$

where a_n denotes the leading coefficient of $f(X)$.

The above version of the Fundamental Theorem of Algebra follows from the following seemingly simpler result by induction on n. Note here that by a nonconstant polynomial one means a polynomial of degree ≥ 1.

Proposition 2.3 (Basic Version of the Fundamental Theorem of Algebra). *A nonconstant polynomial in $\mathbb{C}[X]$ has at least one root in $\mathbb{C}$.*

We will not prove this, but only remark that many different proofs are known, and almost any postgraduate course in mathematics (complex analysis, galois theory, real analysis, topology, ...) can include a proof using the ideas developed in that course.

A field K is said to be *algebraically closed* if every nonconstant polynomial in $K[X]$ has a root in K. This implies that a nonzero polynomial of degree n in $K[X]$ will have n roots, counted with multiplicities. Thus, the Fundamental Theorem of Algebra says that $\mathbb{C}$ is an algebraically closed field, whereas the earlier examples show that $\mathbb{Q}$ and $\mathbb{R}$ are not algebraically closed. It is possible to show that every field is a subfield of an algebraically closed field and the "smallest" algebraically closed field containing K as a subfield is called the *algebraic closure* of K.

Let K be any field. We say that $f(X) \in K[X]$ is *irreducible* if $f(X)$ is a nonconstant polynomial and the only polynomials in $K[X]$ that divide $f(X)$ are constant polynomials or constant multiples of $f(X)$. Now one has the following weak version of the Fundamental Theorem of Algebra that is valid over any field and is in fact, much easier to prove. It shows that if, in addition to allowing multiplicities, we are willing to allow "roots" of higher degrees, then a polynomial of degree n in $K[X]$ always has n roots in K.

Proposition 2.4 (Unique Factorization Theorem). *Let K be a field and $f(X)$ be a nonzero polynomial of degree n in $K[X]$. Then there are distinct monic irreducible polynomials $p_1(X), \ldots, p_r(X) \in K[X]$ and positive integers $e_1, \ldots, e_r$ such that if a_n denotes the leading coefficient of $f(X)$, then*

$$f(X) = a_n p_1(X)^{e_1} \cdots p_r(X)^{e_r}.$$

In particular, if f_i denotes the degree of $p_i(X)$ for $i = 1, \ldots, r$, then

$$e_1 f_1 + \cdots + e_r f_r = n.$$

Polynomials in more than one variable can be considered in an analogous manner. For example, $X^2 - Y^2 + XY$ and $Y^2 - X^3$ are polynomials in 2 variables X and Y. In general, a polynomial in two variables is a finite sum of terms of the form $a_{ij}X^iY^j$, where $a_{ij} \in K$ and i, j are nonnegative integers; the degree of such a term is $i+j$ provided $a_{ij} \neq 0$. The *(total) degree* of a polynomial is the highest among the degrees of the terms appearing in it. The set of all polynomials in two variables X and Y with coefficients in a field K is denoted by $K[X, Y]$. Polynomials in n variables and the notion of degree are defined similarly. The set of all polynomials in n variables $X_1, \ldots X_n$ with coefficients in a field K is denoted by $K[X_1, \ldots, X_n]$. A polynomial in $K[X_1, \ldots, X_n]$ is said to be *homogeneous* if each term has the same degree. For example, $X^2 - Y^2 + XY$ is homogeneous (of degree 2), while $Y^2 - X^3$ and $Y^3 - XY - X - 1$ are not homogeneous. In general, every polynomial can be uniquely expressed as a sum of homogeneous polynomials. More precisely, given any $f(X, Y) \in K[X, Y]$, we can write

$$f(X, Y) = f_0(X, Y) + f_1(X, Y) + \cdots + f_n(X, Y)$$

where either $f_i(X, Y) = 0$ or $f_i(X, Y)$ is a nonzero homogeneous polynomial of degree i. Moreover $f_n(X, Y)$ is nonzero if and only if the degree of $f(X, Y)$ is n. The polynomials $f_i(X, Y)$ are uniquely determined by $f(X, Y)$ and are called the *homogeneous components* of $f(X, Y)$.

A nonhomogeneous polynomial can be converted to a homogeneous polynomial (of the same degree) by introducing a new variable and using it to make every term have the same degree. This process is known as *homogenization*. For example, the homogenization of $Y^2 - X^3$ is $Y^2Z - X^3$, while the homogenization of $Y^3 - XY - X - 1$ is $Y^3 - XYZ - XZ^2 - Z^3$.

In general, if $\deg f(X,Y) = n$ and $f_i(X,Y)$ for $i = 0, 1, \ldots, n$ denote its homogeneous components, then a homogenization of $f(X,Y)$ is given by

$$F(X,Y,Z) = f_0(X,Y)Z^n + f_1(X,Y)Z^{n-1} + \cdots + f_{n-1}(X,Y)Z + f_n(X,Y).$$

We can retrieve the original polynomial $f(X,Y)$ from $F(X,Y,Z)$ by putting $Z = 1$, and this is called *dehomogenization* with respect to the variable Z.

Using dehomogenization and homogenization, we can see that homogeneous polynomials in two variables behave like polynomials in one variable. In particular, we have the following consequence of the Fundamental Theorem of Algebra.

Corollary 2.5. *Let K be an algebraically closed field and let $F(X,Y)$ be a homogeneous polynomial of degree n in $K[X,Y]$. Then there are $(\alpha_1, \beta_1), \ldots, (\alpha_r, \beta_r)$ in $K^2 \setminus \{(0,0)\}$ and positive integers $e_1, \ldots, e_r$ such that*

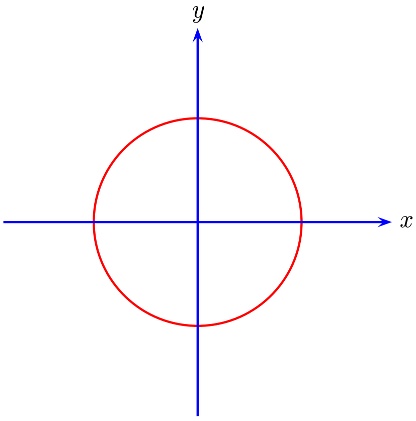
$$F(X,Y) = (\alpha_1 X + \beta_1 Y)^{e_1} \cdots (\alpha_r X + \beta_r Y)^{e_r} \quad \text{with} \quad e_1 + \cdots + e_r = n.$$

Moreover, $(\alpha_1, \beta_1), \ldots, (\alpha_r, \beta_r)$ can be so chosen that no two are proportional to each other, that is, $(\alpha_i, \beta_i) \neq \lambda(\alpha_j, \beta_j)$ for any $\lambda \in K$ and any $i \neq j$.

If the field is not algebraically closed, one can still define the notion of an irreducible polynomial in a similar manner, and show that every nonzero polynomial in $K[X_1, \ldots, X_n]$ can be factored as a product of irreducible polynomials, and moreover, the factorization is unique up to multiplication by nonzero constants and a permutation of the factors. This is sometimes expressed by saying that the ring $K[X_1, \ldots, X_n]$ is a unique factorization domain, or in short, a UFD.

Of course, polynomials in more than one variable do not have finitely many zeros. For example, the zero set of $Y^2 - X$ in K^2 consists of all points of the form (t^2, t) as t varies over K. The zero sets, in general, define curves and in the next section, we will discuss these objects in greater detail.

NOTES AND SUGGESTIONS FOR FURTHER READING: *Older books on algebra such as Chrystal [7] or Burnside and Panton [6] contain a wealth of information about the theory of equations, and it may still be profitable to study them. For a more modern, but friendly and elementary introduction, see Stillwell [30] and Childs [10]. For a more comprehensive account of algebra, one may consult the books of Lang [21] and Abhyankar [4]. For a beautiful generalization of the $\sum e_i f_i = n$ formula in the Unique Factorization Theorem, see Chapter V of Zariski-Samuel [31] or the Kiel notes [15]. To gain a perspective on classical and modern algebra, see the EMS volume edited by Kostrikin and written by Shafarevich [19]; this is a book filled with deep insights and could serve as an excellent bedtime reading for anyone interested in algebra.*



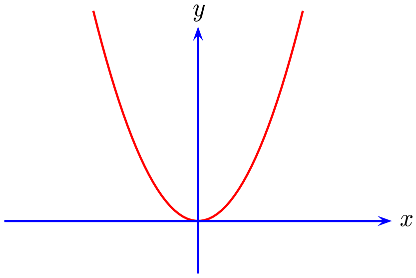


Figure 1: The circle $X^2 + Y^2 - 1 = 0$ and the parabola $Y - X^2 = 0$

3 Analytic Geometry

Analytic geometry, often studied as a precursor to or in conjunction with calculus, consists of studying geometric configurations by means of algebraic equations. For example, we study the circle by means of an equation such as $X^2 + Y^2 - 1 = 0$. In general, we study conic sections that are given by a general equation of degree 2, that is, an equation of the form

$$aX^2 + hXY + bY^2 + gX + fY + c = 0,$$

where a, b, c, f, g, h are constants (say in $\mathbb{R}$) and $(a, h, b) \neq (0, 0, 0)$. We learn to classify conics. Thus, we show that every conic is one of the following: circle, ellipse, parabola, hyperbola or a pair of lines. Here we allow the possibility that the pair of lines may be parallel or may consist of a single coincidental line. To say that a conic is given by a pair of lines effectively means that the defining polynomial is factors into two linear factors. Otherwise, the conic is given by an irreducible polynomial of degree 2.

It may be remarked that the irreducible conics themselves can be classified in two broad categories. The first consisting of circles, ellipses and hyperbolas, and the second consisting of parabolas. If one is working over $\mathbb{C}$, there isn't much difference within the conics in the first category. To see the distinguishing features of the two categories, note that that a circles, ellipses and hyperbolas can be parametrized by rational functions. For example, the circle $X^2 + Y^2 - 1 = 0$ and the hyperbola $XY - 1 = 0$ are

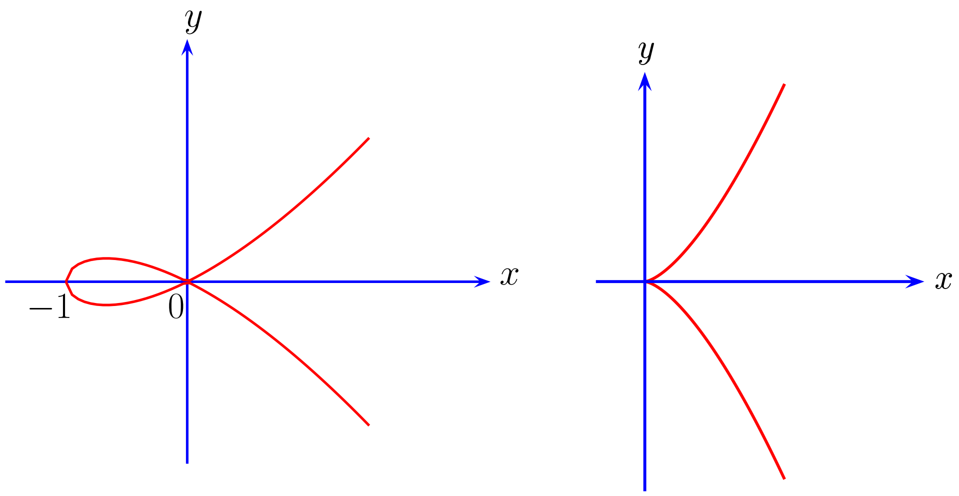


Figure 2: The nodal cubic $Y^2 - X^2 - X^3 = 0$ and the cusp $Y^2 - X^3 = 0$

parametrized by

$$\begin{cases} X = \dfrac{1-t^2}{1+t^2} \\ \\ Y = \dfrac{2t}{1+t^2} \end{cases} \quad \text{and} \quad \begin{cases} X = t \\ \\ Y = \dfrac{1}{t}, \end{cases}$$

respectively. In the case of parabolas, for example, the parabola $Y - X^2 = 0$, there is not only a rational parametrization, but, in fact, a polynomial parametrization, given by $X = t$ and $Y = t^2$.

Having studied conics, one can go on to study cubic curves, such as the nodal cubic (the "alpha curve") defined by $Y^2 - X^2 - X^3 = 0$ or the cuspidal cubic defined by $Y^2 - X^3 = 0$ or the cubic $Y^2 - X^3 + 1 = 0$. [See Figure 3.] In this case it may be noted that the nodal cubic as well as the cuspidal cubic admit a rational parametrization, in fact, a polynomial parametrization, given by

$$\begin{cases} X = (t^2 - 1) \\ \\ Y = t(t^2 - 1) \end{cases} \quad \text{and} \quad \begin{cases} X = t^2 \\ \\ Y = t^3, \end{cases}$$

respectively. For the cubic curve given by $Y^2 - X^3 + 1 = 0$, it can be shown that there is no rational parametrization. [Try to prove this!]

More generally, one can study plane curves such as those given by $f(X, Y) = 0$, where $f(X, Y)$ is a polynomial in two variables. If $f(X, Y)$

is an irreducible polynomial, then the corresponding plane curve is said to be *irreducible* or devoid of multiple components. In general, the irreducible factors of $f(X,Y)$ give rise to the irreducible components of the plane curve given by $f(X,Y)=0$. The *degree* of the plane curve defined by $f(X,Y)$ is simply the total degree of the polynomial $f(X,Y)$.

One of the things studied in analytic geometry is the notion of tangents at a point of a plane curve. For example, let us determine the tangent of the parabola $Y-X^2=0$ at the point $(1,1)$. Most of us would want to compute $\frac{dy}{dx}$, evaluate it at $(1,1)$ to get the slope m and say that the tangent is given by $Y-1=m(X-1)$. Now what about a tangent to $Y-X^{17}+27X^{13}-151X^{11}+71X^2-27X=0$ at the origin or a tangent to $Y^2-X+2=0$ at the point $(2,0)$. In the first example, computing $\frac{dy}{dx}$ is complicated but can be done. In the second, $\frac{dy}{dx}$ can't be computed at $y=0$, but one can still determine the tangent by looking at $\frac{dx}{dy}$. However, what if we consider the curve $Y^2-X^2-X^3=0$ at the origin. Now, calculus fails to give any result. Indeed, those familiar with calculus will say that the tangent to $f(x,y)=0$ at (x_0,y_0) is given by

$$(x-x_0)f_x(x_0,y_0)+(y-y_0)f_y(x_0,y_0)=0.$$

This is fine as long as at $f_x(x_0,y_0)\neq 0$ or $f_y(x_0,y_0)\neq 0$. Indeed, in the former case, $\frac{dx}{dy}$ exists and is given by $-f_y/f_x$, whereas in the latter case, $\frac{dy}{dx}$ exists and is given by $-f_x/f_y$. [Remember the Implicit Function Theorem!] But when $f_x(x_0,y_0)=f_y(x_0,y_0)=0$, the above equation is not helpful. This is what happens in the case of $Y^2-X^2-X^3=0$ at the origin.

Calculus notwithstanding, tangents to any plane curve can be easily computed at any point on the curve. This can be simply done as follows. Given a point (x_0,y_0) on the plane curve $f(X,Y)=0$, make a simple change of coordinates to translate (x_0,y_0) to the origin. In other words, consider $f(X+x_0,Y+y_0)$ and write this as a sum of homogeneous polynomials in X and Y:

$$f(X+x_0,Y+y_0)=f_m(X,Y)+f_{m+1}(X,Y)+\cdots+f_n(X,Y),$$

where $f_i(X,Y)$ is either the zero polynomial or a nonzero homogeneous polyomial of degree i. Moreover, m,n are positive integers so chosen that f_m and f_n are nonzero. If we can factor the *initial form* $f_m(X,Y)$ as a product of non-proportional homogeneous linear factors as

$$f_m(X,Y)=(\alpha_1X+\beta_1Y)^{e_1}\cdots(\alpha_rX+\beta_rY)^{e_r}\quad\text{with}\quad e_1+\cdots+e_r=m,$$

then the lines given by $\alpha_iX+\beta_iY=0$, for $i=1,\dots,r$, are the *tangents* to the curve $f(X,Y)=0$ at (x_0,y_0). For example, using this recipe, we quickly see that the tangent to $Y - X^{17} + 27X^{13} - 151X^{11} + 71X^2 - 27X = 0$ at the origin is the line $Y - 27X = 0$ and the tangent to $Y^2 - X + 2 = 0$ at the point $(2, 0)$ is the line $X = 0$. The curve $Y^2 - X^2 - X^3 = 0$ has two tangents at the origin given by the two factors of $Y^2 - X^2$, namely, the lines given by $Y - X = 0$ and $Y + X = 0$. A factorization of the initial form into powers of homogeneous linear factors is always possible if K is algbraically closed, thanks to Corollary 2.5. Otherwise, one may have to consider tangents of "higher degrees".

NOTES AND SUGGESTIONS FOR FURTHER READING: *Some of the classical texts on the material in this section are Askwith [5], Coolidge [8], Semple and Roth [28], and Salmon [24, 25, 26]. For more on tangents from the point of view of calculus, see [17], and for more on the algebraic viewpoint adopted here, see the "Engineering book" of Abhyankar [3]. A nice and readable analysis of the differences betweeen hyperobolas and parabolas is given in the* Intelligencer *article of Abhyankar [2]. A proof that the cubic* $Y^2 - X^3 + 1 = 0$ *does not admit a rational paramterization can be found, for example, in the book of Reid [23].*

4 Affine Varieties and Hilbert's Nullstellensatz

We have already made a headstart in trying to explain what algebraic geometry is. Indeed, as Abhyankar [3] says, algebraic geometry, at least in its classical form, is an amalgamation of analytic geometry and the theory of equations. The basic objects of study in algebraic geometry are plane curves and more generally, geometric configurations given by the zero-sets of polynomial equations in two or more variables. To make this sound a little more formal, let us introduce some terminology.

Let K be a field. The space $K^n = \{(\alpha_1, \ldots, \alpha_n) : \alpha_1, \ldots, \alpha_n \in K\}$ of all n-tuples of elements of K is called the *affine n-space* over K and denoted by $\mathbb{A}^n_K$ or simply, $\mathbb{A}^n$ when the reference to K is understood from the context. Given any subset S of $K[X_1, \ldots, X_n]$, the set

$$V(S) := \{(\alpha_1, \ldots, \alpha_n) \in \mathbb{A}^n : f(\alpha_1, \ldots, \alpha_n) = 0 \text{ for all } f(X_1, \ldots, X_n) \in S\}$$

of common zeros in $\mathbb{A}^n$ of polynomials in S is called the affine variety defined by S. In general, a subset V of $\mathbb{A}^n$ is called an *affine variety* if $V = V(S)$ for some $S \subseteq K[X_1, \ldots, X_n]$. An affine variety can be defined by several different subsets of $K[X_1, \ldots, X_n]$. For example, if $f, g \in K[X_1, \ldots, X_n]$, then $V(\{f, g\}) = V(\{f + g, f - g\}) = V(\{uf + vg : u, v \in K[X_1, \ldots, X_n]\})$.

In general, for any $S \subseteq K[X_1, \ldots, X_n]$, if $I = \langle S \rangle$ denotes the ideal generated by S in $K[X_1, \ldots, X_n]$, that is, if I denotes the set of all polynomials of the form $u_1 f_1 + \cdots + u_r f_r$, where $f_1, \ldots, f_r$ vary over finite collections of elements of S and $u_1, \ldots, u_r$ vary over $K[X_1, \ldots, X_n]$, then we have $V(S) = V(I)$. With this in view, one may define an affine varieties in $\mathbb{A}^n_K$ as subsets of $\mathbb{A}^n_K$ of the form $V(I)$ where I is an ideal of $K[X_1, \ldots, X_n]$.

By passing to ideals, we easily see that any affine variety (except the one defined by the zero ideal) can be realized as the set of common zeros of infinitely many polynomials. On the other hand, it is more natural to ask whether any affine variety can be defined by common zeros of *finitely many* polynomials. In algebraic setting this corresponds to asking whether every ideal I of $K[X_1, \ldots, X_n]$ is finitely generated. Remarkably, the answer is yes, thanks to the following basic result of Hilbert.

Proposition 4.1 (Hilbert Basis Theorem). *Let K be any field and V be any affine variety in $\mathbb{A}^n_K$. Then there are finitely many polynomials $f_1, \ldots, f_r$ in $K[X_1, \ldots, X_n]$ such that $V = V(\{f_1, \ldots, f_r\})$.*

Now let us consider another natural question. To begin with, note that if I is the unit ideal, that is, if $I = K[X_1, \ldots, X_n]$, then $V(I)$ is clearly the empty set. But what about the converse? In other words, if I is not the unit ideal, then does $V(I)$ have at least one point? Equivalently, if $\{f_i\}$ is a set of polynomials in $K[X_1, \ldots, X_n]$ such that 1 can not be written as a finite linear combination of the f_i's (with coefficients in $K[X_1, \ldots, X_n]$), then do the f_i's have a common zero. In a special case, we have already considered this question. Indeed, suppose $n = 1$ and write $X_1 = X$. Consider an ideal I of $K[X]$ generated by a single polynomial $f(X)$, that is, $I = \{u(X)f(X) : u(X) \in K[X]\}$. [It can be shown that every ideal of $K[X]$ is of this form. Check!] Now I is the unit ideal if and only if $f(X)$ is a nonzero constant polynomial. We have seen that if $K = \mathbb{C}$ or in general, if K is algebraically closed, then a nonconstant polynomial in $K[X]$ has a root in K. This shows that if I is a nonunit ideal of $K[X]$ and K is algebraically closed, then the polynomials in I have a common zero in $\mathbb{A}^1_K$. Thus, the following result, known as *Hilbert's Nullstellensatz* (which means *Hilbert's Zero Point Theorem*), may be viewed as a multidimensional generalization of the Fundamental Theorem of Algebra.

Proposition 4.2 (Hilbert's Nullstellensatz). *Let K be an algebraically closed field. If I is a nonunit ideal of $K[X_1, \ldots, X_n]$, then $V(I)$ is nonempty.*

We have seen that an ideal of $K[X_1, \ldots, X_n]$ gives rise to an affine variety in $\mathbb{A}^n_K$. We may ask whether an affine variety V in $\mathbb{A}^n_K$ determines the corresponding ideal in $K[X_1, \ldots, X_n]$. It is clear that in its simplest form, the answer to this question is No, even when K is algebraically closed. For example, if I_1 and I_2 are ideals in $K[X, Y]$ generated by $X^2 - Y$ and $X^4 - 2X^2Y + Y^2$ respectively, then it is clear that $I_1 \neq I_2$, but $V(I_1) = V(I_2)$, where the latter follows by noting that $X^4 - 2X^2Y + Y^2 = (X^2 - Y)^2$. However, Hilbert's Nullstellensatz shows that this is essentially the only way in which the two ideals can differ. More precisely, we have the following.

Proposition 4.3. *Let K be an algebraically closed field, and let I_1, I_2 be ideals of $K[X_1, \ldots, X_n]$ such that $V(I_1) = V(I_2)$, then for every $f \in I_1$, there is $r \geq 1$ such that $f^r \in I_2$. Likewise, for every $g \in I_2$, there is $s \geq 1$ such that $g^s \in I_1$.*

The above result can be stated in a succinct form using the notion of the radical of an ideal. If I is an ideal of $K[X_1, \ldots, X_n]$, then the set

$$\sqrt{I} := \{f \in K[X_1, \ldots, X_n] : f^m \in I \text{ for some } m \geq 1\}$$

is easily seen to be an ideal of $K[X_1, \ldots, X_n]$ and it is called the *radical* of I. Clearly, $\sqrt{I}$ contains I. One says that I is a *radical ideal* if $\sqrt{I} = I$.

Proposition 4.4. *Let K be an algebraically closed field. If I_1, I_2 are ideals of $K[X_1, \ldots, X_n]$ such that $V(I_1) = V(I_2)$, then $\sqrt{I_1} = \sqrt{I_2}$. As a consequence, there is a one-to-one correspondence between the affine varieties in $\mathbb{A}^n_K$ and the radical ideals in $K[X_1, \ldots, X_n]$.*

The above result may be used to reduce the study of affine algebraic varieties over algebraically closed fields to that of radical ideals in polynomial rings. One may also pass to quotients of polynomial rings by such ideals, which give rise to a class of rings called finitely generated reduced K-algebras. Such rings are studied in courses on commutative algebra.

NOTES AND SUGGESTIONS FOR FURTHER READING: *Proofs of the Hilbert basis theorem and Hilbert's Nullstellensatz can be found in any book on commutative algebra. For a quick introduction, see the AFS notes [16]. For more on affine varieties (and also for a proof of Hilbert's Nullstellensatz), one may consult the books of Kunz [20] or Reid [23].*

5 Projective Varieties and Bezout's Theorem

Now, let us go back to analytic geometry and consider intersections of plane curves. Let us first note that the Fundamental Theorem of Algebra solves completely the problem of intersecting the X-axis (given by $Y =$

0) with curves of the form $Y - f(X) = 0$, where $f(X)$ is a nonconstant polynomial in one variable X with coefficients in $\mathbb{C}$ (or more generally, in an algebraically closed field K). The points of intersection are of the form $(\alpha, 0)$, where α is a root of $f(X)$. Moreover, if the degree of $f(X)$ is n or equivalently, if the degree of the plane curve given by $Y - f(X)$ is n, then there are exactly n points of intersection provided we count $(\alpha, 0)$ m times if the multiplicity of α as a root of $f(X)$ is m. We can do the same thing in a little more general case. Suppose C is any plane curve given by $f(X, Y) = 0$ and L is a line given parametrically by

$$\begin{cases} X = c_1 + d_1 t \\ Y = c_2 + d_2 t \end{cases}$$

where $c_1, d_1, c_2, d_2 \in K$ with $(d_1, d_2) \neq (0, 0)$. The points of intersection of C with L correspond to the roots of the polynomial $\phi(t) := f(c_1 + d_1 t, c_2 + d_2 t)$ in one variable t with coefficients in K. If $\deg f(X, Y) = n$ and $f_0, f_1, \ldots, f_n$ denote the homogeneous components of $f(X, Y)$, then

$$\phi(t) = f(c_1 + d_1 t, c_2 + d_2 t) = f_n(d_1, d_2) t^n + \text{ terms of lower degree in } t.$$

By Corollary 2.5, $f_n(X, Y)$ has at most finitely many non-proportional roots, and thus most values of (d_1, d_2) will satisfy $f_n(d_1, d_2) \neq 0$. In this case, by the Fundamental Theorem of Algebra, $\phi(t)$ has roots $t_1, \ldots, t_r \in K$ with respective multiplicities $e_1, \ldots, e_r$ satisfying $e_1 + \cdots + e_r = n$. Hence if $P_i := (c_1 + d_1 t_i, c_2 + d_2 t_i)$ for $i = 1, \ldots, r$, then $P_1, \ldots, P_r$ are in $C \cap L$ and moreover if we define the *intersection multiplicity* of C and L at P_i to be

$$I(C, L; P_i) := e_i = \text{the multiplicity of } t_i \text{ as a root of } \phi(t) = f(c_1 + d_1 t, c_2 + d_2 t),$$

then we have

$$\sum_{i=1}^{r} I(C, L; P_i) = n = (\deg C)(\deg L).$$

In other words, the curve C of degree n and the line L of degree 1 meet in exactly n points, counted properly. But what happens in the special case when $f_n(d_1, d_2) = 0$? To understand this, let us consider a very simple example. Suppose $f(X, Y) = X - Y + 1$ and the line L is given parametrically by $X = t$ and $Y = t$. Now $\phi(t) = f(t, t) = 1$ and this clearly has no roots. Basically, C and L are two parallel lines and they have no point in common in the affine plane $\mathbb{A}^2$. It would be nice if we can somehow count properly so that such exceptional cases do not arise, and we are able to say that a two lines always meet in a point. Intuition comes to rescue here. If we think of two parallel lines as railway tracks, then we "see" that they seem to meet at "infinity". So it would be nice if we can postulate a point at infinity in each parallel direction and regard that any two parallel lines meet in the corresponding point at infinity. This leads to an extension of the affine plane $\mathbb{A}^2_K$ to the *projective plane* $\mathbb{P}^2_K$, which may be defined as follows.

Consider the set $K^3 \setminus \{(0,0,0)\}$ of all triples of elements of K, except the zero triple. Define an equivalence relation $\sim$ on this set as follows.

$$(x,y,z) \sim (x',y',z') \iff x' = \lambda x, y' = \lambda y \text{ and } z' = \lambda z \text{ for some } \lambda \in K.$$

In other words, two triples are equivalent if they are proportional to each other. Define $\mathbb{P}^2_K$ as the set of all equivalence classes. Elements of $\mathbb{P}^2_K$ may be written as $[x : y : z]$ where $x, y, z \in K$ and not all x, y, z are zero, and it should be understood that $[x : y : z] = [\lambda x : \lambda y : \lambda z]$ for any $\lambda \in K$ with $\lambda \neq 0$. Clearly, the points of $\mathbb{P}^2_K$ fall in two categories: those $[x : y : z]$ for which $z \neq 0$ and those for which $z = 0$. The former can be represented by $[a : b : 1]$ for a uniquely determined $(a,b) \in \mathbb{A}^2_K$, while the latter are of the form $[\alpha : \beta : 0]$ for some $\alpha, \beta \in K$, not both zero. Thus

$$\mathbb{P}^2_K = \mathbb{A}^2_K \cup L_\infty \quad \text{where} \quad L_\infty = \{[\alpha : \beta : 0] : (\alpha,\beta) \in K^2 \setminus \{(0,0)\}\}.$$

The elements of L_∞ may be thought of as the points at infinity that are added to the affine plane $\mathbb{A}^2_K$ to obtain the projective plane $\mathbb{P}^2_K$; indeed, $[\alpha : \beta : 0]$ corresponds to the point at infinity in the direction of the line $\alpha X + \beta Y = 0$ (or for that matter, the line $\alpha X + \beta Y = \gamma$ for any $\gamma \in K$).

Now, how does one think of a curve in the projective plane? That's easy. While an affine plane curve is given by an equation of the form $f(X,Y) = 0$, where $f(X,Y)$ is a polynomial in $K[X,Y]$, a *projective plane curve* is given by an equation of the form $F(X,Y,Z) = 0$, where $F(X,Y,Z)$ is a homogeneous polynomial in $K[X,Y,Z]$. Notice that if $F(X,Y,Z) \in K[X,Y,Z]$ is homogeneous of degree n, then for any $x,y,z \in K$ and any $\lambda \in K$, we have $F(\lambda x, \lambda y, \lambda z) = \lambda^n F(x,y,z)$. Thus, $F(x,y,z) = 0$ if and only if $F(\lambda x, \lambda y, \lambda z) = 0$ for all $\lambda \in K$. Hence it makes sense to talk about the zeros of $F(X,Y,Z)$ in $\mathbb{P}^2_K$. In case $F(X,Y,Z)$ is obtained from homogenizing $f(X,Y) \in K[X,Y]$, then the zero-set of $F(X,Y,Z)$ in $\mathbb{P}^2_K$ consist of the zero-set of $f(X,Y)$ in $\mathbb{A}^2_K$ (identified as a subset of $\mathbb{P}^2_K$ as given above) together with a finite number of points in L_∞; these are the *points at infinity* of the curve C given by $f(X,Y) = 0$. To see this, note that if

$$f(X,Y) = f_0(X,Y) + f_1(X,Y) + \cdots + f_n(X,Y) \quad \text{where} \quad f_n(X,Y) \neq 0,$$

is the decomposition of $f(X,Y)$ into homogeneous components, then the homogenization $F(X,Y,Z)$ is given by

$$F(X,Y,Z) = f_0(X,Y)Z^n + f_1(X,Y)Z^{n-1} + \cdots + f_{n-1}(X,Y)Z + f_n(X,Y).$$

Clearly, $F(a, b, 1) = f(a, b)$, while $F(a, b, 0) = f_n(a, b)$. By Corollary 2.5, $f_n(X, Y)$ has at most n non-proportional homogeneous linear factors (and exactly n if we count properly). Thus, a curve of degree n has at most n points at infinity and these are given by the "zeros" of the degree form $f_n(X, Y)$. One can think of the projective plane $\mathbb{P}^2_K$ as the *geometric completion* of the affine plane $\mathbb{A}^2_K$ and the projective plane curve $F(X, Y, Z) = 0$ as *geometric completion* of the affine plane curve $f(X, Y) = 0$.

Now, just as solving a single quadratic equation ($X^2 + 1 = 0$) yielded a miracle in the form of the Fundamental Theorem of Algebra, arranging that any two lines always meet in a point yields a miracle known as Bezout's Theorem. Before stating this, we note that if two projective plane curves over an algebraically closed field have a common component, that is, if the corresponding homogeneous polynomials $F(X, Y, Z)$ and $G(X, Y, Z)$ have a nonconstant homogeneous common factor, then they clearly meet in infinitely many points.

Proposition 5.1 (Bezout's Theorem). *Let K be an algebraically closed field. If C and D are projective plane curves of degrees m and n respectively and if C and D do not have a common component, then C and D meet in exactly mn points, counted properly. More precisely, $C \cap D$ is a finite subset of $\mathbb{P}^2_K$ containing most mn points, and moreover,*

$$\sum_{P \in C \cap D} I(C, D; P) = mn = (\deg C)(\deg D),$$

where $I(C, D; P)$ denotes the "intersection multiplicity" of C and D at P.

To be sure, we have not given a precise definition of the intersection multiplicity of two curves at a point of their intersection. We have of course done this when one of the curves is a line. It is not difficult to see that a similar definition can be given for the intersection of a projective plane curve with a line in $\mathbb{P}^2_K$ or more generally, a curve given by nice parametric equations. Here is one of the several possible definitions for those who know some algebra. It may be noted that since the intersection multiplicity is a local notion, one may assume that the point is at "finite distance" by making a suitable change of coordinates, if necessary. Thus, suppose C and D are affine plane curves given by polynomials $f(X, Y)$ and $g(X, Y)$ in $K[X, Y]$. Let $P = (a, b) \in \mathbb{A}^2_K$ be a point of intersection. Consider the quotient field $K(X, Y)$ of $K[X, Y]$ (which consists of rational functions of the form $u(X, Y)/v(X, Y)$, where $u(X, Y), v(X, Y) \in K[X, Y]$ with $v(X, Y) \neq 0$) and the subring $\mathcal{O}_P$ of $K(X, Y)$ consisting of those quotients $u(X, Y)/v(X, Y)$ for which $v(a, b) \neq 0$. Look at the ideal $I = \langle f, g \rangle \mathcal{O}_P$

generated by $f(X, Y)$ and $g(X, Y)$ in the ring $\mathcal{O}_P$. If C and D have no common components, that is if $f(X, Y)$ and $g(X, Y)$ have no nonconstant common factor, then it can be shown that $\mathcal{O}_P/I$ is a finite-dimensional vector space over the field K. One defines $I(C, D : P) = \dim_K \mathcal{O}_P/I$.

NOTES AND SUGGESTIONS FOR FURTHER READING: *For a more detailed introduction, filled with examples, heuristics and applications, to the projective plane, projective plane curves and Bezout's Theorem, see the "Engineering book" of Abhyankar [4] and also his Chauvnet prize winning article [1] in the* MONTHLY. *The latter also contains several different definitions of intersection multiplicity. A complete proof of Bezout's Theorem using the definition given at the end of this section can be found, for example, in the appendix to Silverman and Tate [29] or in Fulton's book [14].*

6 Epilogue

These notes (which, in fact, contain a little more material than what was actually covered in the lecture) barely scratch the surface of algebraic geometry. There are important strands that we have left untouched. One such is the connection between algebraic curves and compact Riemann surfaces. This is a line of development that starts from the problem of determination of the arc-length of an ellipse and goes on to a host of interesting and important topics. We refer to the first few pages of Abhyankar's MONTHLY article [1] followed by the China lectures of Griffits [18] for an introduction to this topic. For a different viewpoint, see Chevalley [9]. Number theoretic aspects can be found, for example in the book of Silverman and Tate [29]. For a wholesome and substantive introduction to algebraic geometry, the books of Mumford [22] and Shafarevich [27] are recommended. The latter contains an appendix outlining the historical development of algebraic geometry. See also the *Final Comments* in the book of Reid [23] for a view of the history and sociology of algebraic geometry. A number of applications of algebraic geometry, both within and outside mathematics, can be found in the book of Cox, Little and O'Shea [11].

The reader may have noticed that although these notes were once meant for the participants of a course on Advances in Control Theory, there has been no mention thus far of Control Theory. The reason is obvious. Namely, the author has no expertise on Control Theory. However, the two volumes of Falb [12] may be useful for anyone wishing to understand the relation between Algebraic Geometry and Control Theory.

References

[1] S. S. Abhyankar, *Historical Ramblings in Algebraic Geometry and Related Algebra*, American Math. Monthly, **83** (1976), 409–448.

[2] S. S. Abhyankar, *What is the difference between a parabola and a hyperbola?*, Math. Intelligencer, **10** (1988), 36–43.

[3] S. S. Abhyankar, *Algebraic Geometry for Scientists and Engineers*, American Math. Society, 1990.

[4] S. S. Abhyankar, *Lecture on Algebra*, Vol. I, World Scientific, 2006.

[5] E. H. Askwith, *A Course of Pure Geometry*, Cambridge University Press, 1937.

[6] W. S. Burnside and A. W. Panton, *Theory of Equations*, Vol. I and II, Dover, 1960.

[7] G. Chrystal, *Algebra*, Parts I and II, 6th ed., Chelsea, 1959.

[8] J. L. Coolidge, *A Treatise on Algebraic Plane Curves*, Dover, 1959.

[9] C. Chevalley, *Introduction to the Theory of Algebraic Functions of One Variable*, American Mathematical Society, 1951.

[10] L. N. Childs, *A Concrete Introduction to Higher Algebra*, Springer-Verlag, 1979.

[11] D. Cox, J. Little and D. O'Shea, *Using Algebraic Geometry*, 2nd Ed., Springer, 2005.

[12] P. Falb, *Methods of Algebraic Geometry in Control Theory*, Part 1: *Scalar Linear Systems and Affine Algebraic Geometry*; Part 2: *Multivariable Linear Systems and Projective Algebraic Geometry*, Birkhäuser, 1990.

[13] B. Fine and G. Rosenberger, *The Fundamental Theorem of Algebra*, Springer–Verlag, 1997.

[14] W. Fulton, *Algebraic Curves*, Addison-Wesley, 1969.

[15] S. R. Ghorpade, *Lectures on Topics in Algebraic Number Theory*, Universität zu Kiel, Kiel, Germany, 2002. [Available: `http://www.math.iitb.ac.in/~srg/Lecnotes.html`]

[16] S. R. Ghorpade, *Lectures on Commutative Algebra*, NBHM Annual Foundational School (AFS-II), Pune, 2006. [Available: `http://www.math.iitb.ac.in/~srg/Lecnotes.html`]

[17] S. R. Ghorpade and B. V. Limaye, *A Course in Calculus and Real Analysis*, Springer, 2006. [Information about low price Indian edition as well as a dynamic errata is available at: `http://www.math.iitb.ac.in/~srg/acicara/`]

[18] P. A. Griffiths, *Introduction to Algebraic Curves*, American Mathematical Society, 1989.

[19] A. I. Kostrikin and I. R. Shafarevich (Eds), *Algebra I: Basic Notions of Algebra*, Springer–Verlag, 1990.

[20] E. Kunz, *Introduction to Commutative Algebra and Algebraic Geometry*, Birkhäuser, 1985.

[21] S. Lang, *Algebra*, 2nd ed., Addison-Wesley, 1984.

[22] D. Mumford, *Algebraic Geometry I: Complex Projective Varieties*, Springer-Verlag, 1976.

[23] M. Reid, *Undergraduate Algebraic Geometry*, Cambridge University Press, 1988.

[24] G. Salmon, *A Treatise on Conic Sections*, 6th Ed., Longmans, Green and Co., 1879.

[25] G. Salmon, *A Treatise on Higher Plane Curves*, 3rd Ed., Chelsea, 1960.

[26] G. Salmon, *A Treatise on Analytic Geometry in Three Dimensions*, Vol. I and II, 7th Ed. and 5th Ed., Chelsea, 1958 and 1965.

[27] I. R. Shafarevich, *Basic Algebraic Geometry*, Vol. 1 and 2, 2nd Ed., Springer, 1994.

[28] J. G. Semple and L. Roth, *Introduction to Algebraic Geometry*, Clarendon Press, Oxford, 1949.

[29] J. H. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer, 1992.

[30] J. Stillwell, *Elements of Algebra*, Springer, 1994.

[31] O. Zariski and P. Samuel, *Commutative Algebra*, Vol. 1, Van Nostrand, 1958.