VIII. Background for Algebraic Geometry, 447-519

DOI: 10.3792/euclid/9781429799928-8

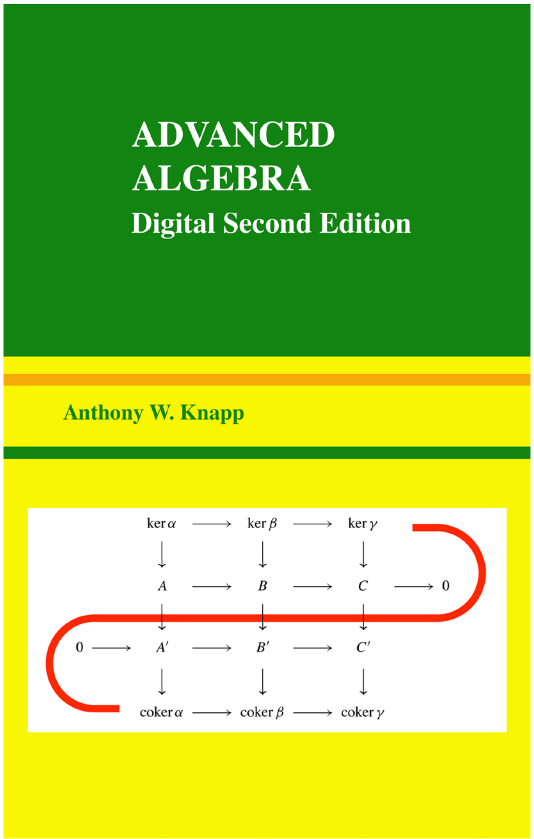
from

Advanced Algebra

Digital Second Edition

Anthony W. Knapp

Full Book DOI: 10.3792/euclid/9781429799928
ISBN: 978-1-4297-9992-8



ADVANCED
ALGEBRA
Digital Second Edition

Anthony W. Knapp

Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733–1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/∼aknapp

Title: Advanced Algebra
Cover: Content of the Snake Lemma; see page 185.

Mathematics Subject Classification (2010): 11–01, 13–01, 14–01, 16–01, 18G99, 55U99, 11R04, 11S15, 12F99, 14A05, 14H05, 12Y05, 14A10, 14Q99.

First Edition, ISBN-13 978-0-8176-4522-9
©2007 Anthony W. Knapp
Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN
©2016 Anthony W. Knapp
Published by the Author

CHAPTER VIII

Background for Algebraic Geometry

Abstract. This chapter introduces aspects of the algebraic theory of systems of polynomial equations in several variables.

Section 1 gives a brief history of the subject, treating it as one of two early sources of questions to be addressed in algebraic geometry.

Section 2 introduces the resultant as a tool for eliminating one of the variables in a system of two such equations. A first form of Bezout's Theorem is an application, saying that if $f(X, Y)$ and $g(X, Y)$ are polynomials of respective degrees m and n whose locus of common zeros has more than mn points, then f and g have a nontrivial common factor. This version of the theorem may be regarded as pertaining to a pair of affine plane curves.

Section 3 passes to projective plane curves, which are nonconstant homogeneous polynomials in three variables, two such being regarded as the same if they are multiples of one another. Versions of the resultant and Bezout's Theorem are valid in this context, and two projective plane curves defined over an algebraically closed field always have a common zero.

Sections 4–5 introduce intersection multiplicity for projective plane curves. Section 4 treats a line and a curve, and Section 5 treats the general case of two curves. The theory in Section 4 is completely elementary, and a version of Bezout's Theorem is proved that says that a line and a curve of degree d have exactly d common zeros, provided the underlying field is algebraically closed, the zeros are counted as often as their intersection multiplicities, and the line does not divide the curve. Section 5 makes more serious use of algebraic background, particularly localizations and the Nullstellensatz. It gives an indication that ostensibly simple phenomena in the subject can require sophisticated tools to analyze.

Section 6 proves a version of Bezout's Theorem appropriate for the context of Section 5: if F and G are two projective plane curves of respective degrees m and n over an algebraically closed field, then either they have a nontrivial common factor or they have exactly mn common zeros when the intersection multiplicities of the zeros are taken into account.

Sections 7–10 concern Gröbner bases, which are finite generating sets of a special kind for ideals in a polynomial algebra over a field. Section 7 sets the stage, introducing monomial orders and defining Gröbner bases. Section 8 establishes a several-variable analog of the division algorithm for polynomials in one variable and derives from it a usable criterion for a finite set of generators to be a Gröbner basis. From this it is easy to give a constructive proof of the existence of Gröbner bases and to obtain as consequences solutions of the ideal-membership problem and the proper-ideal problem. Section 9 obtains a uniqueness theorem under the condition that the Gröbner basis be reduced. Adjusting a Gröbner basis to make it reduced is an easy matter. A consequence of the uniqueness result is a solution of the ideal-equality problem. Section 10 gives two theorems concerning solutions of systems of polynomial equations. The Elimination Theorem identifies in terms of Gröbner bases those members of the ideal that depend only on a certain subset of the variables. The Extension Theorem, proved under the additional assumption that the underlying field is algebraically closed,

gives conditions under which a solution to the subsystem of equations that depend on all but one variable can be extended to a solution of the whole system. The latter theorem makes use of the theory of resultants.

1. Historical Origins and Overview

Modern algebraic geometry grew out of early attempts to solve simultaneous polynomial equations in several variables and out of the theory of Riemann surfaces. We shall discuss the first of these sources in the present chapter and the second of the sources in Chapter IX.

Serious consideration of simultaneous polynomial equations of degree > 2 dates to a 1750 book[1] by Gabriel Cramer (1704–1752), who may be better known for Cramer's rule in connection with determinants. Cramer was interested in various aspects of the zero loci of polynomials in two variables with real coefficients. Thinking of the zero locus, we refer to a nonconstant polynomial in two variables as a plane curve.

One of the problems of interest to Cramer was to find the number of points in the plane that would uniquely determine a plane curve of degree n up to a constant multiple. Cramer gave the answer $\frac{1}{2}n(n+3)$ to this problem. For example, when $n = 2$, if we normalize matters by taking the coefficient of x^2 to be 1, then the possible quadratic polynomials

$$f(x, y) = x^2 + bxy + cy^2 + dx + ey + f$$

involve five unknown coefficients. Each condition $f(x_i, y_i) = 0$ gives a linear condition on the coefficients, and Cramer was able to write down explicitly a plane curve through the given points in question by introducing determinants and applying his rule to solve the problem.

Already with this much description the reader will see a certain subtlety—that there will be special choices of the five points for which existence or uniqueness will fail. We could also ask about the effect of multiplicities: what does it mean geometrically to take two or more of the points to be equal, and how does such an occurrence affect the number of points that can be specified?

Cramer noticed a subtlety that is less easy to resolve, even in hindsight. If we are given any two plane curves of degree 3, then Cardan's formula says that we can solve one equation for y in terms of x, obtaining three expressions in x; then we can substitute for y in the other equation each of the three expressions in x and obtain a cubic equation in x each time. In other words, we should expect up to 9 points of intersection for two cubics, and 9 should sometimes occur. (The various

[1] G. Cramer, *Introduction à l'Analyse des Lignes courbes algébriques*, Chez les Frères Cramer & Cl. Philibert, Geneva, 1750.

forms of Bezout's Theorem, which came a little later, confirm this argument.) The number of points that determine a cubic completely is $\frac{1}{2}n(n+3)$ for $n = 3$, i.e., is 9. Thus we have 9 points determining a unique cubic, and yet the second cubic goes through these 9 points as well. What is happening? This question has come to be known as **Cramer's paradox**.

Explaining this kind of mystery became an early impetus for the development of algebraic geometry.

The question of the number of points of intersection had been the subject of conjecture for some time earlier, and it was expected that two plane curves of respective total degrees m and n in some sense had mn points of intersection. Étienne Bezout (1730–1783) took up this question and dealt with parts of it rigorously. The quadratic case can be solved by finding one variable in terms of the other and by substituting, but let us handle it by the method that Bezout used. If we view each polynomial as quadratic in y and having coefficients that depend on x, then we have a system

$$\begin{aligned} a_0 + a_1 y + a_2 y^2 &= 0, \\ b_0 + b_1 y + b_2 y^2 &= 0. \end{aligned}$$

Instead of regarding this as a system of two equations for y, we regard it as a system of two homogeneous linear equations for variables x_0, x_1, x_2, where $x_0 = 1, x_1 = y, x_2 = y^2$. We can get two further equations by multiplying each equation by y:

$$\begin{aligned} a_0 y + a_1 y^2 + a_2 y^3 &= 0, \\ b_0 y + b_1 y^2 + b_2 y^3 &= 0, \end{aligned}$$

and then we have four homogeneous linear equations for $x_0 = 1, x_1 = y, x_2 = y^2, x_3 = y^3$. Since the system has the nonzero solution $(1, y, y^2, y^3)$, the determinant of the coefficient matrix must be 0. Remembering that the coefficients depend on x, we see that we have eliminated the variable y and obtained a polynomial equation for x without using any solution formula for polynomials in one variable. The device that Bezout introduced for this purpose—the determinant of the coefficient matrix—is called the **resultant** of the system and is a fundamental tool in handling simultaneous polynomial equations. With it Bezout went on in 1779 to give a rigorous proof that when two polynomials in (x, y) are set equal to 0 simultaneously, one of degree m and the other of degree n, then there cannot be more than mn solutions unless the two polynomials have a common factor. This is a first form of Bezout's Theorem and is proved in Section 2.

In order to have a chance of obtaining a full complement of mn solutions, we make three adjustments—allow complex solutions instead of just real solutions (even in the case $(m, n) = (2, 1)$), consider "projective plane curves" instead of ordinary plane curves to allow for solutions at infinity (even in the case $(m, n) =$

$(1, 1)$), and introduce a suitable notion of intersection number of two plane curves at a point in order to take multiplicities into account (even in the case $(m, n) = (2, 1)$). We shall allow complex solutions already in Section 2, and we shall make an adjustment for projective plane curves in Section 3. The issue of intersection multiplicity is more complicated. The beginnings of a classical approach to it are indicated in Section 4, and a somewhat more modern approach appears in Section 5. With the full theory of intersection multiplicities of projective plane curves in place, we obtain a general form of Bezout's Theorem[2] in Section 6.

The theory of the resultant can be extended in various ways, but we shall largely not pursue this matter. Studies of zero loci of systems of equations took a more geometric turn in the first part of the nineteenth century through the work of Julius Plücker (1801–1858) and others, but these matters will be left for an implicit discussion in Chapter X. Instead, we skip to a development that began with the doctoral thesis of Bruno Buchberger in 1965. Buchberger was interested in being able to decide when a polynomial is a member of an ideal that is specified by a finite list of generators. For this purpose he learned that each ideal has a special finite set of generators that is unique once certain declarations are made. He devised an algorithm for determining such a set of generators,[3] and he gave the name "Gröbner basis" to the set, in honor of his thesis advisor.[4] The special unique such basis is called a "reduced Gröbner basis."

An unfortunate feature of the algorithm (and even of later improved algorithms) is that Gröbner bases are extraordinarily complicated to calculate. The timing of Buchberger's discovery was therefore especially fortuitous, coming when computers were becoming more common, more economical, and more powerful.

Buchberger was able to give a test for membership in an ideal in terms of a multivariable division algorithm involving any Gröbner basis. Other general problems involving ideals were solvable as well. Because of the uniqueness of the reduced Gröbner basis, two ideals are identical if and only if their reduced Gröbner bases are equal. When some of the theory of resultants was incorporated into the theory of Gröbner bases, these bases could also be used to address various questions of identifying zero loci. Other problems involving ideals could be addressed by similar methods. The theory has flowered tremendously since its initial discovery and by the present day has found many imaginative applications to applied problems. Sections 7–10 give an introductory account of this important theory.

[2] A correct proof of the general form of the theorem seems to have been published for the first time by Georges-Henri Halphen (1844–1889) in 1873.

[3] Devising the algorithm was Buchberger's real contribution, since the abstract existence of the special set of generators is an easy consequence of the Hilbert Basis Theorem and had already been used in papers of H. Hironaka in 1964.

[4] Wolfgang Gröbner (1899–1980). The name is often spelled out as "Groebner," particularly when it is used in connection with computer algorithms.

2. Resultant and Bezout's Theorem

Let A be a unique factorization domain. The case that $A = K[X_1, \dots, X_r]$ for a field K will be the main case of interest for us. If f and g are polynomials in $A[X]$ of the form

$$f(X) = f_0 + f_1X + \cdots + f_mX^m,$$
$$g(X) = g_0 + g_1X + \cdots + g_nX^n,$$

with m and n both positive, then we let $\mathcal{R}(f, g)$ be the $(m+n)$-by-$(m+n)$ matrix

$$\begin{pmatrix}
f_0 & f_1 & \cdots & f_{m-1} & f_m & 0 & 0 & 0 & \cdots & 0 \\
0 & f_0 & \cdots & f_{m-2} & f_{m-1} & f_m & 0 & 0 & \cdots & 0 \\
\vdots & & & \ddots & & & & \ddots & & \vdots \\
0 & \cdots & & & & f_0 & & & \cdots & f_m \\
g_0 & g_1 & \cdots & & & g_{n-1} & g_n & 0 & \cdots & 0 \\
0 & g_0 & \cdots & & & g_{n-2} & g_{n-1} & g_n & \cdots & 0 \\
\vdots & & \ddots & & & & & & \ddots & \vdots \\
0 & \cdots & & g_0 & g_1 & & \cdots & & & g_n
\end{pmatrix},$$

in which there are n rows above the g_0 in the first column and there are m remaining rows. The **resultant** of f and g is the determinant

$$R(f, g) = \det \mathcal{R}(f, g).$$

Theorem 8.1. If A is a unique factorization domain and if f and g are nonzero members of $A[X]$ of the form $f(X) = \sum_{i=0}^{m} f_iX^i$ and $g(X) = \sum_{j=0}^{n} g_jX^j$ with $m > 0$ and $n > 0$ and with at least one of f_m and g_n nonzero, then the following are equivalent:

(a) f and g have a common factor of degree > 0 in X,

(b) $af + bg = 0$ for some nonzero a and b in $A[X]$ with $\deg a < n$ and $\deg b < m$.

(c) $R(f, g) = 0$.

Regard $R(f, g)$ as a constant polynomial in X. When $R(f, g) \neq 0$, there exist unique a and b in $A[X]$ such that $a(X)f(X) + b(X)g(X) = R(f, g)$ with $\deg a < n$ and $\deg b < m$. Both the polynomials a and b are nonzero if both $f(X)$ and $g(X)$ are nonconstant.

REMARKS. The theorem says that $af + bg = R(f, g)$ holds in every case for which at least one of the coefficients f_m and g_n is nonzero. Sometimes the theorem appears in texts with the assumption that both coefficients are nonzero; in this connection, see Problem 5 at the end of the chapter. When $R(f, g) = 0$, the theorem does not point to a useful way to identify a common factor; the division algorithm can be used for this purpose in some circumstances, but the use of Gröbner bases as in Section 7 will be more helpful.

PROOF. Let us prove the equivalence of (a) and (b). Suppose that (a) holds. If u is a nonconstant polynomial in X that divides both f and g, let us write $f = bu$ and $g = -au$. Then $af + bg = 0$. Also, $\deg a + \deg u = \deg g$; since $\deg u > 0$, $\deg a < \deg g \leq n$. Similarly $\deg b < m$. Thus (b) holds. Conversely suppose that (b) holds, so that $af = -bg$ with a and b nonzero and with $\deg a < n$ and $\deg b < m$. Suppose that $f_m \neq 0$. The equality $af = -bg$ shows that f divides bg. Since $\deg b < m = \deg f$, f cannot divide b. But $A[X]$ is a unique factorization domain, and thus there is some prime factor p of f of positive degree such that p^k for some k divides f but not b. Then p divides both f and g, and (a) holds. A similar argument works if $g_n \neq 0$.

Now we prove the equivalence of (b) and (c). Let F be the field of fractions of A. We set up a one-one correspondence between polynomials $a(X)$ in $A[X]$ of degree at most $n-1$ and n-dimensional row vectors $(\,\alpha_0 \quad \alpha_1 \quad \cdots \quad \alpha_{n-1}\,)$ with entries in A by the formula

$$a(X) = \alpha_0 + \alpha_1 X + \cdots + \alpha_{n-1} X^{n-1},$$

and similarly we set up one-one correspondences for degrees at most $m-1$ and at most $m+n-1$ by the formulas

$$\begin{aligned} b(X) &= \beta_0 + \beta_1 X + \cdots + \beta_{m-1} X^{m-1}, \\ c(X) &= \gamma_0 + \gamma_1 X + \cdots + \gamma_{m+n-1} X^{m+n-1}. \end{aligned}$$

Examining the form of $\mathcal{R}(f, g)$, we see that the matrix equality

$$\begin{aligned} (\,\alpha_0 \quad \alpha_1 \quad \cdots \quad \alpha_{n-1} \quad \beta_0 \quad \cdots \quad \beta_{m-1}\,)\, &\mathcal{R}(f, g) \\ &= (\,\gamma_0 \quad \gamma_1 \quad \cdots \quad \gamma_{m+n-1}\,) \qquad (*) \end{aligned}$$

holds if and only if the polynomial equality

$$a(X)f(X) + b(X)g(X) = c(X). \qquad (**)$$

holds. If (b) holds, then $af = -bg$, and $(**)$ shows that $c = 0$. That is, $(\,\gamma_0 \quad \gamma_1 \quad \cdots \quad \gamma_{m+n-1}\,)$ is the 0 row vector. Interpreting $(*)$ as a matrix equality over F and assuming that a and b are not both 0, we see that the transpose of $\mathcal{R}(f, g)$ has a nontrivial null space. Therefore $R(f, g) = \det \mathcal{R}(f, g) = 0$. This proves (c). Conversely if (c) holds, then we can find row vectors $(\,\alpha_0 \quad \alpha_1 \quad \cdots \quad \alpha_{n-1}\,)$ and $(\,\beta_0 \quad \beta_1 \quad \cdots \quad \beta_{m-1}\,)$ not both 0, having entries in F, such that the left side of $(*)$ equals the 0 row vector. Clearing fractions, we may assume that $(\,\alpha_0 \quad \alpha_1 \quad \cdots \quad \alpha_{n-1}\,)$ and $(\,\beta_0 \quad \beta_1 \quad \cdots \quad \beta_{m-1}\,)$ have entries in A. Referring to $(*)$, we obtain $af + bg = 0$ with $\deg a$ at most $n-1$ and $\deg b$ at most $m-1$. We know that at least one of a and b is nonzero, and we have to see that both are nonzero. The situation is symmetric in a and b. If a were to equal 0, then we would have $bg = 0$ and we could conclude that $b = 0$ because $g \neq 0$. So we would obtain the contradiction $a = b = 0$. This proves (b).

For the last statement of the theorem, suppose that $R(f, g) \neq 0$. Then Cramer's rule applied over the field of fractions F of A shows that the matrix inverse of $\mathcal{R}(f, g)$ is of the form

$$\mathcal{R}(f, g)^{-1} = R(f, g)^{-1}\mathcal{S}(f, g),$$

where $\mathcal{S}(f, g)$ is a matrix with entries in A. Consequently the row vector

$$(\, R(f, g) \quad 0 \quad \cdots \quad 0\,)\,\mathcal{R}(f, g)^{-1}$$

has entries in A, and we can define members $\alpha_0, \dots, \alpha_{n-1}, \beta_0, \dots, \beta_{m-1}$ of A by

$$\begin{aligned}(\,\alpha_0 \quad \alpha_1 \quad \cdots \quad \alpha_{n-1} \quad \beta_0 \quad \cdots \quad \beta_{m-1}\,)& \\ = (\, R(f, g) \quad 0 \quad \cdots \quad 0\,)\,\mathcal{R}(f, g)^{-1}.&\end{aligned}$$

Then $(*)$ holds with $(\,\gamma_0 \quad \gamma_1 \quad \cdots \quad \gamma_{m+n-1}\,) = (\, R(f, g) \quad 0 \quad \cdots \quad 0\,)$, and the equality $(**)$ shows that $a(X)f(X) + b(X)g(X) = R(f, g)$. If both f and g are nonconstant, then neither $a(X)$ nor $b(X)$ can be 0, since otherwise the equation would show that $R(f, g)$ is a nonconstant polynomial. □

Theorem 8.2 (Bezout's Theorem). Let K be any field, and let $f(X, Y)$ and $g(X, Y)$ be nonconstant polynomials in $K[X, Y]$, of exact respective degrees m and n. If the locus of common zeros of f and g in K^2 has more than mn points, then f and g have a nonconstant common factor in $K[X, Y]$.

PROOF. For most of the proof, we assume that K is infinite. Arguing by contradiction, suppose that f and g both vanish at distinct points (x_i, y_i) for $1 \leq i \leq mn + 1$, and suppose that f and g have no nonconstant common factor. Since there are only finitely many members c of K such that $y_i - y_j = c(x_i - x_j)$ for some i and j with $i \neq j$ and since K is assumed to be infinite, we can find c in K such that $y_i - y_j \neq c(x_i - x_j)$ for all i and j with $i \neq j$. For this c, $y_i - cx_i \neq y_j - cx_j$ when $i \neq j$, and therefore the second coordinates of the points $(x_i, y_i - cx_i)$ are distinct. The common zeros of $f(X, Y)$ and $g(X, Y)$ include the points (x_i, y_i), and thus the common zeros of $f(X, Y + cX)$ and $g(X, Y + cX)$ include the $mn + 1$ points $(x_i, y_i - cx_i)$ whose second coordinates are distinct.

In other words, there is no loss of generality in assuming that the given polynomials f and g vanish at $mn + 1$ points whose second coordinates are distinct. Regard $f(X, Y)$ and $g(X, Y)$ as members $f(X)$ and $g(X)$ of $A[X]$, where $A = K[Y]$, and write

$$f(X) = f_0 + f_1 X + \cdots + f_{m'} X^{m'},$$
$$g(X) = g_0 + g_1 X + \cdots + g_{n'} X^{n'},$$

with each f_i and g_i in A and with $f_{m'} \neq 0$ and $g_{n'} \neq 0$. Here $m' \leq m$ and $n' \leq n$.

Let us rule out the possibility that $m' = 0$ or $n' = 0$. Indeed, if we had $m' = 0$, then the polynomial f would be nonzero and would depend on Y alone. Since f is nonzero and has degree $m \geq 1$, it has at most m roots. But we are assuming that f and g vanish at $mn + 1$ points whose Y coordinates are distinct, and the inequalities $m \leq mn < mn + 1$ therefore give a contradiction. Thus $m' \neq 0$. Similarly $n' \neq 0$. So Theorem 8.1 is applicable.

Form the square matrix $\mathcal{R}(f, g)$ of size $m' + n'$ and its determinant $R(f, g)$. The latter is a member of $K[Y]$, and Theorem 8.1 shows that it cannot be 0, since f and g are assumed to have no nonconstant common factor in $K[X, Y]$.

Let us bound the degree of the member $R(f, g) = \det \mathcal{R}(f, g)$ of $K[Y]$. Each term in the expansion of the determinant is of the form

$$\pm \prod_{1 \leq i \leq m'+n'} \mathcal{R}(f, g)_{i,\sigma(i)} \tag{$*$}$$

for some permutation σ of $\{1, \ldots, m' + n'\}$. Here $\mathcal{R}(f, g)_{ij}$ is given by

$$\mathcal{R}(f, g)_{ij} = \begin{cases} f_{j-i} & \text{for } 1 \leq i \leq n' \text{ and for } j \text{ with } i \leq j \leq m' + i, \\ 0 & \text{for } 1 \leq i \leq n' \text{ and for all other } j, \\ g_{j+n'-i} & \text{for } n' + 1 \leq i \leq n' + m' \text{ and for } j \\ & \text{with } i \leq n' + j \leq m' + i, \\ 0 & \text{for } n' + 1 \leq i \leq n' + m' \text{ and for all other } j. \end{cases}$$

In addition, the degree of f_{j-i} as a member of $K[Y]$ is at most $m - (j - i)$, and the degree of $g_{j+n'-i}$ is at most $n - (j + n' - i) = (n - n') + (i - j)$. Setting $j = \sigma(i)$, we see that the degree of $(*)$ is at most

$$\sum_{1 \leq i \leq n'} (m - \sigma(i) + i) + \sum_{n'+1 \leq i \leq m'+n'} ((n - n') + (i - \sigma(i)))$$
$$= mn' + m'(n - n') = mn - (m - m')(n - n') \leq mn.$$

Thus $R(f, g)$ is a nonzero polynomial in $K[Y]$ of degree at most mn. Consequently it has at most mn roots.

Theorem 8.1 shows that $af + bg = R(f, g)$ for suitable members a and b of $K[X, Y]$. Recalling that f and g are assumed to vanish at $mn + 1$ points whose second coordinates are distinct, we see that $R(f, g)$ vanishes at each of these second coordinates, and we arrive at a contradiction.

Now we can allow K to be finite. Let K' be an infinite extension. We have just seen that f and g have a nonconstant factor in $K'[X, Y]$. Without loss of generality, this factor depends nontrivially on X. Theorem 8.1 applied with $A = K'[Y]$ shows that $R[f, g] = 0$. The same theorem with $A = K[Y]$ then shows that f and g have a common factor in $A[X] = K[X, Y]$ depending nontrivially on X. $\square$

Let us introduce some geometric language for the situation in Theorem 8.2. **Affine n-space** over a field K is the set of n-dimensional column vectors

$$\mathbb{A}^n = \mathbb{A}^n_{K_{\text{alg}}} = \big\{(x_1, \dots, x_n) \in K^n_{\text{alg}}\big\}$$

with entries in a fixed algebraic closure K_{alg} of K. The set of K **rational points**, or K **points**, in $\mathbb{A}^n$ is the subset

$$\mathbb{A}^n_K = \big\{(x_1, \dots, x_n) \in K^n\big\}.$$

We shall comment on the appearance of K_{alg} in these definitions shortly.

Members of $\mathbb{A}^n$ are called **points** in n-dimensional affine space, and the functions $P \mapsto x_j(P)$ give the **coordinates** of the points. If L is any field between K and K_{alg}, then any polynomial f in $K[X_1, \dots, X_n]$ defines a corresponding polynomial function from $\mathbb{A}^n_L$ into L.

For algebraic geometry the case of interest for Sections 1–6 of this chapter is the case $n = 2$. The way of viewing a curve is influenced by Cramer's thinking as discussed in Section 1: the particular polynomial that defines a curve is important, not just the zero locus in the affine plane, but two curves are to be regarded as the same if each is a nonzero multiple of the other. We can incorporate this viewpoint into algebraic language by defining an **affine plane curve** C over the field K to be any nonzero proper principal ideal[5] in $K[X, Y]$. The curve is an **affine plane line** if the degree of any generator is 1.

In practice in studying affine plane curves, there is ordinarily no need to distinguish between a polynomial and the principal ideal that it generates, and we shall feel free to refer to an affine plane curve $C = (f)$ as f when there is no possibility of confusion.

The zero locus of a curve is the corresponding geometric notion, but it can readily be empty, as is the case with $X^2 + Y^2 + 1$ when $K = \mathbb{R}$. On the other hand, the Nullstellensatz (Theorem 7.1) ensures that the zero locus will be nonempty if the underlying field is algebraically closed. Thus we define the **zero locus** $V(C) = V((f))$ of the curve $C = (f(X, Y))$ by[6]

[5] *Warning:* This definition will be changed slightly in Chapter IX and again in Chapter X to reflect changed emphasis in those chapters.

[6] The letter "V" is the letter that is commonly used in the notation for a zero locus. It stands for "variety," a notion that we have not yet defined. But beware: not all objects labeled with a "V" are actually varieties the way the term is normally defined. An affine plane curve will turn out to be a variety exactly when the generating polynomial f is prime in $K_{\text{alg}}[X, Y]$.

$$V(C) = V_{K_{\rm alg}}(C) = \big\{(x, y) \in K_{\rm alg}^2 \mid f(x, y) = 0\big\}.$$

This is the same as the set of all (x, y) such that every member of the ideal C vanishes at (x, y). The set of K **rational points**, or K **points**, of C is

$$V_K(C) = V_K((f)) = \big\{(x, y) \in K \mid f(x, y) = 0\big\}.$$

When we are content to refer to an affine curve $C = (f)$ as f, we are content also to write $V(f)$ in place of $V(C) = V((f))$.

In Chapter X, under the assumption that K is algebraically closed, we shall extend these definitions from the case $n = 2$ and C as above to the case that n is general and C is replaced by any ideal I in $K[x_1, \dots, X_n]$. The set $V(I)$ of common zeros of the members of I in $K^n = K_{\rm alg}^n$ will be called an "affine algebraic set." The case of affine n-space itself arises when the ideal is 0.

For general K, not necessarily algebraically closed, it is meaningful to consider the set $V_K(I)$ of K rational points, i.e., the subset of common zeros lying in K^n. For $I = 0$ and $V(I) = \mathbb{A}^n$, the distinction between $V_K(I)$ and $V_{K_{\rm alg}}(I)$ is hardly worth mentioning, but the distinction is well worth making for general I and is made for the case $V(I) = \mathbb{A}^n$ for consistency. Although the study of sets $V_K(I)$ is of importance in number theory, in geometry over $\mathbb{R}$, and in other areas, we shall not pursue it in Chapter X for lack of space.

Returning to Theorem 8.2, we see that the statement concerns $V_K(C) \cap V_K(D)$, where C and D are the principal ideals $C = (f)$ and $D = (g)$ in $K[X, Y]$. The theorem says that if $V_K(C) \cap V_K(D)$ contains more than mn points, then there is a nonzero principal ideal h with $(h) \subseteq (f) \cap (g)$.

3. Projective Plane Curves

Section 2 dealt with intersections of affine plane curves. Even over an algebraically closed field, two affine plane curves need not intersect. An example is the pair of straight lines $X + Y - 1$ and $X + Y - 2$, whose locus of common zeros is empty. To get these lines to intersect, we have to introduce "points at infinity." The projective plane is the device for including such points.

Let K be a field, and let $K_{\rm alg}$ be an algebraic closure. The **projective plane** over K is defined set theoretically as the quotient of $K_{\rm alg}^3 - \{0\}$ by an equivalence relation:

$$\mathbb{P}^2 = \mathbb{P}^2_{K_{\rm alg}} = \big\{(x, y, w) \in K_{\rm alg}^3 - \{0\}\big\}\big/\sim,$$

where $(x', y', w') \sim (x, y, w)$ if $(x', y', w') = \lambda(x, y, w)$ for some $\lambda \in K_{\rm alg}^\times$. The set of K **rational points**, or K **points**, of $\mathbb{P}^2$ is the quotient

$$\mathbb{P}^2_K = \big\{(x, y, w) \in K^3 - \{0\}\big\}\big/\sim,$$

where $(x', y', w') \sim (x, y, w)$ if $(x', y', w') = \lambda(x, y, w)$ for some $\lambda \in K$. When there is a need to be careful, we shall write $[x, y, w]$ for the member of $\mathbb{P}^2_K$ corresponding to (x, y, w) in $K^3 - \{0\}$. But often there will not be such a need, and we shall simply refer to (x, y, w) as a member of $\mathbb{P}^2_K$. Both $\mathbb{P}^2$ and $\mathbb{P}^2_K$ have additional structure on them, given by "affine local coordinates," and we come to that matter later in this section.

Let us record briefly the obvious generalization of the projective plane to other dimensions: **Projective n-space** over K is defined set theoretically as the quotient

$$\mathbb{P}^n = \mathbb{P}^n_{K_{\text{alg}}} = \big\{(x_1, \dots, x_{n+1}) \in K^{n+1}_{\text{alg}} - \{0\}\big\}\big/\sim,$$

where $(x'_1, \dots, x'_{n+1}) \sim (x_1, \dots, x_{n+1})$ if $(x'_1, \dots, x'_{n+1}) = \lambda(x_1, \dots, x_{n+1})$ for some $\lambda \in K^\times_{\text{alg}}$. The set $\mathbb{P}^n_K$ of K **rational points** of $\mathbb{P}^n$ is the set defined in similar fashion using just nonzero vectors in K^{n+1} and scalars in $K^\times$.

Scalar-valued functions on $\mathbb{P}^n_K$ are of little interest because they amount to scalar-valued functions of $K^n - \{0\}$ that are unchanged when $(x_1, \dots, x_n)$ is replaced by a multiple of itself. A polynomial of this kind, for example, is necessarily constant. Instead, the polynomials of interest that are related to $\mathbb{P}^n_K$ are "homogeneous polynomials." A **monomial** in $K[X_1, \dots, X_{n+1}]$ is a polynomial of the form $X_1^{j_1} \cdots X_{n+1}^{j_{n+1}}$; its **total degree** is $\sum_{i=1}^{n+1} j_i$. We say that a nonzero F in $K[X_1, \dots, X_{n+1}]$ is **homogeneous** of degree $d \geq 0$ if every monomial appearing in F with nonzero coefficient has total degree d. By convention the 0 polynomial is homogeneous of every degree. We write $K[X_1, \dots, X_{n+1}]_d$ for the set of homogeneous polynomials of degree d. Each such F satisfies

$$F(\lambda x_1, \dots, \lambda x_{n+1}) = \lambda^d F(x_1, \dots, x_{n+1})$$

for all $(x_1, \dots, x_{n+1}) \in K^{n+1}$ and $\lambda \in K^\times$. Conversely the fact that the mapping of polynomials into polynomial functions is one-one for an infinite field implies that homogeneous polynomials over an infinite field can be detected by this property.

Let us assemble some further properties of homogeneous polynomials: The monomials of total degree d form a K basis of the vector space $K[X_1, \dots, X_{n+1}]_d$; this fact follows from the definition of polynomials over K. To calculate the dimension of $K[X_1, \dots, X_{n+1}]_d$, consider the problem of taking d factors X on which to place subscripts and using n dividers to separate the X_1's from the X_2's and so on. The number of monomials in question is just the number of ways of selecting the n dividers from among the $d + n$ symbols and dividers. Thus we obtain the important formula

$$\dim_K K[X_1, \dots, X_{n+1}]_d = \binom{d+n}{n}.$$

Lemma 8.3. Any polynomial factor of a homogeneous polynomial over a field K is homogeneous.

PROOF. Write $F = F_1F_2$ nontrivially. Let d_1 and e_1 be the highest and lowest total degrees of terms in F_1, and let d_2 and e_2 be the highest and lowest total degrees of terms in F_2. The product of the terms of total degree d_1 in F_1 and the terms of total degree d_2 in F_2 is nonzero and is the d_1d_2 total-degree part of F. The product of the terms of total degree e_1 in F_1 and the terms of total degree e_2 in F_2 is nonzero and is the e_1e_2 total-degree part of F. Since F is homogeneous, $d_1d_2 = e_1e_2$. It follows that $d_1 = e_1$ and $d_2 = e_2$; thus F_1 and F_2 are homogeneous. □

An ideal I in $K[X_1, \dots, X_{n+1}]$ is called a **homogeneous ideal** if it is the sum over $d \geq 0$ of its intersections with $K[X_1, \dots, X_{n+1}]_d$:

$$I = \bigoplus_{d=0}^{\infty} (I \cap K[X_1, \dots, X_{n+1}]_d).$$

The sum is to be regarded as a direct sum of vector spaces. For such an ideal, we can compute the quotient $K[X_1, \dots, X_{n+1}]/I$ term by term:

$$K[X_1, \dots, X_{n+1}]/I = \bigoplus_{d=0}^{\infty} K[X_1, \dots, X_{n+1}]_d \big/ (I \cap K[X_1, \dots, X_{n+1}]_d).$$

We can often recognize a homogeneous ideal from its generators: an ideal with a set of generators that are all homogeneous is necessarily a homogeneous ideal. In fact, if an ideal I has homogeneous generators F_j, then the most general member of I is a finite sum of terms A_jF_j. The terms of total degree d in A_jF_j are the product of F_j with the terms in A_j of total degree $d - \deg F_j$, and each such term is in I. Hence each member of I is a sum of homogeneous polynomials that lie in I, and the assertion follows.

In the setting of $\mathbb{P}^2$, projective plane curves over K are initially defined to be nonconstant *homogeneous* polynomials in $K[X, Y, W]$. Although such polynomials are not well defined on the projective plane, their zero loci are well defined subsets of $\mathbb{P}^2$. As in the affine case, the particular polynomial that defines a curve is important, not just the zero locus, but two curves are to be regarded as the same if each is a nonzero multiple of the other. We can incorporate this viewpoint into algebraic language by defining a **projective plane curve** of **degree** $d > 0$ over the field K to be any nonzero proper principal ideal in $K[X, Y, W]$ generated by a homogeneous polynomial of degree d. Such an ideal is necessarily homogeneous. In the special cases that $d = 1$, 2, 3, or 4, the curve is called a **projective line**, **conic**, **cubic**, or **quartic** respectively.

Just as in the affine case, in practice in studying projective plane curves, there is often no need to distinguish between a homogeneous polynomial and the homogeneous principal ideal that it generates, and we shall feel free to refer to a projective plane curve $C = (F) \subseteq K[X, Y, W]$ as F when there is no possibility of confusion.

If (F) is a projective plane curve of degree d, then its zero locus is denoted by

$$V((F)) = V_{K_{\mathrm{alg}}}((F)) = \big\{[x, y, w] \in \mathbb{P}^2 \bigm| F(x, y, w) = 0\big\}.$$

The locus

$$V_K((F)) = \big\{[x, y, w] \in \mathbb{P}^2_K \bigm| F(x, y, w) = 0\big\}$$

is called the set of K **rational points**, or K **points**, of the curve. When we allow ourselves to refer to the curve simply as F, then we can write $V(F)$ in place of $V((F))$.

The affine plane $\mathbb{A}^2_K = \{(x, y)\}$ has a standard one-one embedding into the projective plane $\mathbb{P}^2_K$. Namely we map (x, y) into $[x, y, 1]$. The set that is missed by the image is the set with $w = 0$, which is the set of K rational points of the line L with $L(X, Y, W) = W$, a line called the **line at infinity**. We shall denote this line by W. The points of $V_K(W)$, i.e., those with $w = 0$, are called the **points at infinity**.

Except for the line at infinity, lines in $\mathbb{P}^2_K$ correspond under restriction exactly to lines in K^2. Namely the projective line $L(X, Y, W) = aX + bY + cW$ corresponds to the affine line $l(x, y) = aX + bY + c$, and vice versa. In certain ways the geometry of $\mathbb{P}^2_K$ is simpler than the geometry of $\mathbb{A}^2_K$:

(i) Two distinct lines in $\mathbb{P}^2_K$ intersect in a unique point. In fact, we set up the system of equations

$$\begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix} \begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Since the lines are distinct, the coefficient matrix has rank 2. Thus the kernel has dimension 1, and there is just one point $[x, y, w]$ in the intersection.

(ii) Two distinct points in $\mathbb{P}^2_K$ lie on a unique line. In fact, we set up the system of equations

$$\begin{pmatrix} x & y & w \\ x' & y' & w' \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and argue in similar fashion.

Along with the embedding of $\mathbb{A}^2_K$ into $\mathbb{P}^2_K$ is a correspondence between projective curves and affine curves. Let us work with the polynomials themselves, without identifying each polynomial with every nonzero scalar multiple of itself. The passage from a nonzero homogeneous polynomial $F(X, Y, W)$ of degree

$d > 0$ to a polynomial $f(X, Y)$ is given by $f(X, Y) = F(X, Y, 1)$. The mapping $F \mapsto f$ is a substitution homomorphism, and it therefore respects products. However, the degree may drop in the process, and in particular $f(X, Y)$ is a constant if and only if $F(X, Y, W)$ is a multiple of W^d.

In the reverse direction if $f(X, Y)$ is a polynomial of degree e, then $f(X, Y)$ arises from a polynomial $F(X, Y, W)$, but we have to specify the degree d of F and we must have $d \geq e$. Operationally we obtain F by inserting a power of W into each term of f to make the total degree of the term become d. For example, with $f(X, Y) = Y^2 + XY + X^3$ if the desired degree is 3, then $F(X, Y, W) = Y^2W + XYW + X^3$. On the other hand, if the desired degree is 4, then $F(X, Y, W) = Y^2W^2 + XYW^2 + X^3W$.

The formula for this reverse process is $F(X, Y, W) = W^d f(XW^{-1}, YW^{-1})$. That is, F is given by a substitution homomorphism, followed by multiplication by a power of W. From this fact, we can read off conclusions of the following kind:

> If polynomials $f(X, Y)$ and $g(X, Y)$ are obtained from homogeneous polynomials $F(X, Y, W)$ and $G(X, Y, W)$ by taking $W = 1$, then there exist integers r and s such that the polynomial $W^rF(X, Y, W) + W^sG(X, Y, W)$ is homogeneous and such that $f(X, Y) + g(X, Y)$ is obtained from it by taking $W = 1$.

As we mentioned above, $\mathbb{P}^2_K$ has more structure than simply the structure of a set. About any point in $\mathbb{P}^2_K$ we can introduce various systems of "affine local coordinates." The idea is to imitate what happens in the definition of a manifold: the whole manifold is covered by charts, each giving an invertible mapping of a set in the manifold to an open subset of Euclidean space. Here a single system of affine local coordinates plays the role of a chart; it puts $\mathbb{A}^2_K$ into one-one correspondence with the complement of the zero locus of a line in $\mathbb{P}^2_K$.

Let Φ be a member of the matrix group GL(3, K). Then Φ maps the set K^3 of column vectors in one-one fashion onto K^3 and passes to a one-one map of $\mathbb{P}^2_K$ onto $\mathbb{P}^2_K$ called the **projective transformation** corresponding to Φ. Two Φ's give the same map of $\mathbb{P}^2_K$ if and only if they are multiples of one another. The group action of GL(3, K) on $\mathbb{P}^2_K$ is transitive because GL(3, K) acts transitively on $K^3 - \{(0, 0, 0)\}$.

If L is the projective line whose coefficients are given by the row vector $(a \quad b \quad c)$ and if Φ is is in GL(3, K), then the row vector $(a \quad b \quad c)\,\Phi^{-1}$ defines a new projective line L^Φ, and the K rational points of L^Φ are given by

$$V_K(L^\Phi) = \Phi(V_K(L)).$$

In fact, let $\begin{pmatrix} x \\ y \\ w \end{pmatrix}$ be in $V_K(L)$. Then $\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} = \Phi \begin{pmatrix} x \\ y \\ w \end{pmatrix}$ is in $\Phi(V_K(L))$ and satisfies

$$(\,a \quad b \quad c\,)\,\Phi^{-1}\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} = 0;$$

hence it is in $V_K(L^\Phi)$. Conversely if $\left(\begin{smallmatrix} x' \\ y' \\ w' \end{smallmatrix}\right)$ is in $V_K(L^\Phi)$, then $\left(\begin{smallmatrix} x \\ y \\ w \end{smallmatrix}\right) = \Phi^{-1}\left(\begin{smallmatrix} x' \\ y' \\ w' \end{smallmatrix}\right)$ satisfies

$$(\,a \quad b \quad c\,)\begin{pmatrix} x \\ y \\ w \end{pmatrix} = (\,a \quad b \quad c\,)\,\Phi^{-1}\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} = 0,$$

and thus $\left(\begin{smallmatrix} x' \\ y' \\ w' \end{smallmatrix}\right)$ is Φ of something in $V_K(L)$.

To form the analog of a chart, fix $[x_0, y_0, w_0]$ in $\mathbb{P}^2_K$. Choose (by transitivity) some Φ in $\mathrm{GL}(3, K)$ with $\Phi(x_0, y_0, w_0) = (0, 0, 1)$. Then we can define **affine local coordinates** on $\Phi^{-1}(K \times K \times \{1\})$ to K^2 by the one-one map

$$\varphi(\Phi^{-1}(x, y, 1)) = (x, y).$$

This definition generalizes the standard embedding of the affine plane K^2 into $\mathbb{P}^2_K$ earlier; that embedding was the case $\Phi = 1$.

EXAMPLES OF AFFINE LOCAL COORDINATES FOR $\mathbb{P}^2_K$.

(1) Suppose $(x_0, y_0, w_0) = (x_0, y_0, 1)$. We can choose $\Phi = \left(\begin{smallmatrix} 1 & 0 & -x_0 \\ 0 & 1 & -y_0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$. Then

$$\Phi\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -x_0 \\ 0 & 1 & -y_0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x - x_0 \\ y - y_0 \\ 1 \end{pmatrix}.$$

In this case, the local coordinates are defined on

$$\Phi^{-1}(K \times K \times 1) = K \times K \times 1$$

and are given by

$$\begin{aligned} \varphi(x, y, 1) &= \varphi(\Phi^{-1}(\Phi(x, y, 1))) \\ &= \varphi(\Phi^{-1}(x - x_0, y - y_0, 1)) = (x - x_0, y - y_0). \end{aligned}$$

This Φ is handy for reducing behavior about $(x_0, y_0, 1)$ in $\mathbb{P}^2_K$ to behavior about $(0, 0)$ in K^2.

(2) Suppose $(x_0, y_0, w_0) = (0, 1, 0)$. We can choose $\Phi = \left(\begin{smallmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right)$. Then

$$\Phi \begin{pmatrix} x \\ 1 \\ w \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ 1 \\ w \end{pmatrix} = \begin{pmatrix} w \\ x \\ 1 \end{pmatrix}$$

and

$$\varphi(x, 1, w) = \varphi(\Phi^{-1}(\Phi(x, 1, w))) = \varphi(\Phi^{-1}(w, x, 1)) = (w, x).$$

This Φ is handy for studying behavior near one of the points at infinity in $\mathbb{P}^2_K$.

We can use affine local coordinates to examine the behavior of a projective plane curve "near a particular point," by which is meant "with that point as the center point in the analysis." To examine behavior near $(0, 0, 1)$, we use the correspondence $f(X, Y) = F(X, Y, 1)$ that we discussed earlier. For a general point, we make use of the fact that whenever F is a homogeneous polynomial of degree d, then so is $F \circ \Phi^{-1}$. To examine the behavior of F near a point (x_0, y_0, w_0) in $K^3 - \{(0, 0, 0)\}$, we choose Φ in $\mathrm{GL}(3, K)$ with $\Phi(x_0, y_0, w_0) = (0, 0, 1)$, and we define

$$f(X, Y) = F(\Phi^{-1}(X, Y, 1)).$$

Under this correspondence the behavior of F at (x_0, y_0, w_0) is reflected in the behavior of f at $(0, 0)$. We call $f(X, Y)$ the **local expression** for F in the affine local coordinates determined by Φ. This local expression is a polynomial in $K[X, Y]$, and it is nonconstant unless F is a scalar multiple of $(W \circ \Phi)^d$ for some d.

EXAMPLES, CONTINUED.

(1) Suppose that $(x_0, y_0, w_0) = (x_0, y_0, 1)$ and that $\Phi = \left(\begin{smallmatrix} 1 & 0 & -x_0 \\ 0 & 1 & -y_0 \\ 0 & 0 & 1 \end{smallmatrix}\right)$. Computation gives

$$\Phi^{-1} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} x + x_0 \\ y + y_0 \\ 1 \end{pmatrix},$$

and the corresponding local expression for a projective plane curve F is

$$f(X, Y) = F(X + x_0, Y + y_0, 1).$$

For the projective plane curve

$$F(X, Y, W) = X^2Y + XYW + W^3$$

and the same Φ, the local expression $f(X, Y)$ splits into homogeneous terms as

$$\begin{aligned} f(X,Y) = (x_0^2 y_0 + x_0 y_0 + 1) + (x_0^2 Y + 2x_0 y_0 X + x_0 Y + y_0 X) \\ + (y_0 X^2 + 2x_0 XY + XY) + (X^2 Y). \end{aligned}$$

We shall use this splitting in the next section in the first example of intersection multiplicity.

(2) Suppose that $(x_0, y_0, w_0) = (0, 1, 0)$ and that $\Phi = \begin{pmatrix} 0&0&1 \\ 1&0&0 \\ 0&1&0 \end{pmatrix}$. Then

$$\Phi^{-1}\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} y \\ 1 \\ x \end{pmatrix},$$

and the local expression for a projective plane curve F relative to this Φ is

$$f(X,Y) = F(Y, 1, X).$$

For the same projective plane curve F as in Example 1, namely

$$F(X,Y,W) = X^2Y + XYW + W^3,$$

we obtain

$$f(X,Y) = (Y^2 + XY) + (X^3).$$

We shall examine this example further in the next section.

In this way we have associated to each projective plane curve F and to the system of affine local coordinates determined by a member Φ of $\mathrm{GL}(3, K)$ a local expression that is a nonzero polynomial in $K[X, Y]$. Conversely if the degree d and the member Φ of $\mathrm{GL}(3, K)$ are given and if f in $K[X, Y]$ is nonzero of degree at most d, then we can reconstruct a projective plane curve F of degree d whose local expression relative to Φ is f. We have only to form the unique homogeneous polynomial G of degree d with $f(X, Y) = G(X, Y, 1)$ and then put $F = G \circ \Phi$.

With these preparations in place, we return to a consideration of resultants and Bezout's Theorem. Our objective is to rephrase Theorem 8.2 to take advantage of properties of the projective plane.

Lemma 8.4. Let K be a field, let A be the polynomial ring $A = K[x_1, \dots, x_r]$, and let f and g be members of $A[X]$ of the form

$$\begin{aligned} f(X) &= f_0 + f_1 X + \cdots + f_m X^m, \\ g(X) &= g_0 + g_1 X + \cdots + g_n X^n, \end{aligned}$$

where f_j is a member of A homogeneous of degree $m' - j$ and g_j is a member of A homogeneous of degree $n' - j$. Then the resultant $R(f, g)$ is a homogeneous member of A of degree $mn' + m'n - mn$.

REMARKS. In the application to proving Theorem 8.5, we will have $m' = m$ and $n' = n$, and then $R(f, g)$ is homogeneous of degree mn. Problem 8 at the end of the chapter concerns a situation for which $m' \neq m$ and $n' \neq n$.

PROOF. There is no loss of generality in assuming that K is algebraically closed, hence in particular is infinite. Each nonzero entry $\mathcal{R}(f, g)_{ij}$ of $\mathcal{R}(f, g)$ is a coefficient of f or of g. For each entry, define $p(i, j)$ such that $\mathcal{R}(f, g)_{ij}(tx_1, \dots, tx_r) = t^{p(i,j)}\mathcal{R}(f, g)_{ij}(x_1, \dots, x_r)$. The assembled matrix $\mathcal{R}$ with powers of t in place is

$$\begin{pmatrix} t^{m'} f_0 & t^{m'-1} f_1 & \cdots & t^{m'-m} f_m & \cdots \\ 0 & t^{m'} f_0 & \cdots & & \\ \vdots & & \ddots & & \\ t^{n'} g_0 & t^{n'-1} g_1 & \cdots & t^{n'-n} g_n & \cdots \\ 0 & t^{n'} g_0 & \cdots & & \end{pmatrix}. \tag{$*$}$$

It turns out that there is a function $q(i)$ such that $r(j) = q(i) + p(i, j)$ depends only on j. Here $t^{q(i)}$ is the i^{th} entry of

$$(t^{n'}, t^{n'-1}, \dots, t^{n'-n+1}; t^{m'}, t^{m'-1}, \dots, t^{m'-m+1}).$$

The matrix $(*)$ with $t^{q(i)}$ multiplying every entry of the i^{th} row is

$$\begin{pmatrix} t^{n'} t^{m'} f_0 & t^{n'} t^{m'-1} f_1 & \cdots & t^{n'} t^{m'-m} f_m & \cdots \\ 0 & t^{n'-1} t^{m'} f_0 & \cdots & & \\ \vdots & & \ddots & & \\ t^{m'} t^{n'} g_0 & t^{m'} t^{n'-1} g_1 & \cdots & t^{m'} t^{n'-n} g_n & \cdots \\ 0 & t^{m'-1} t^{n'} g_0 & \cdots & & \end{pmatrix}. \tag{$**$}$$

In $(**)$, $t^{r(j)}$ is the j^{th} entry of $(t^{m'+n'}, t^{m'+n'-1}, \dots, t^{m'+n'-m-n+1})$. Then we have

$$t^u R(f, g)(tx_1, \dots, tx_r) = t^v R(f, g)(x_1, \dots, x_r),$$

where $u = \sum_i q(i)$ and $v = \sum_j r(j)$. So

$$R(f, g)(tx_1, \dots, tx_r) = t^{v-u} R(f, g)(x_1, \dots, x_r).$$

In other words, $R(f, g)$ is a homogeneous function. Since K is infinite, $R(f, g)$ is homogeneous as a member of A. Computing u and v, we find that $u = mm' + nn' - \frac{1}{2}m(m-1) - \frac{1}{2}n(n-1)$ and $v = (m+n)(m'+n') - \frac{1}{2}(m+n)(m+n-1)$. Therefore $v - u = mn' + m'n - mn$, and the degree of homogeneity of $R(f, g)$ is $mn' + m'n - mn$. □

Theorem 8.5 (Bezout's Theorem). Let K be a field, let K_{alg} be an algebraic closure, and suppose that F in $K[X, Y, W]_m$ and G in $K[X, Y, W]_n$ are projective plane curves. Then their locus $V(F) \cap V(G)$ of common zeros in $\mathbb{P}^2_{K_{\text{alg}}}$ is nonempty. If this zero locus has more than mn points, then F and G have as a common factor some homogeneous polynomial in (X, Y, W) of positive degree.

REMARKS. For two polynomials $f(X, Y)$ and $g(X, Y)$ in affine space, application of Theorem 8.1 concerning the resultant in the Y variable involves checking that at least one of the polynomials has the expected degree in the Y variable, and doing so may not be so easy. In the projective setting, this problem disappears if we apply a projective transformation and arrange that $[0, 0, 1]$ not be on the zero locus of one of the given polynomials, say $F(X, Y, W)$. In fact, if F is in $K[X, Y, W]_m$, then the coefficient of W^m has to be a constant, and this term is the only term of F that contributes to the value of F at $(0, 0, 1)$. With the above adjustment the coefficient must be nonzero, and Theorem 8.1 is applicable.

PROOF. Without loss of generality, we may assume throughout that K is algebraically closed. Write F and G in the form

$$\begin{aligned} F(X, Y, W) &= f_0 + f_1 W + \cdots + f_m W^m \qquad \text{with } f_j \in K[X, Y]_{m-j}, \\ G(X, Y, W) &= g_0 + g_1 W + \cdots + g_n W^n \qquad \text{with } g_j \in K[X, Y]_{n-j}. \end{aligned} \tag{$*$}$$

Pick a point (x, y, w) at which F is nonzero, and move it to $(0, 0, 1)$ by a projective transformation, so that $F(0, 0, 1) \neq 0$. Regarding F and G as polynomials in W, with coefficients in $A = K[X, Y]$, we form $R(F, G)$, which Lemma 8.4 identifies as a member of $K[X, Y]_{mn}$.

Since $R(F, G)$ is homogeneous as a member of $K[X, Y]$ and since K is algebraically closed, we can choose a point $(x_0, y_0) \neq (0, 0)$ with $R(F, G)(x_0, y_0) = 0$. Then the resultant of $F(x_0, y_0, W)$ and $G(x_0, y_0, W)$ is 0, and Theorem 8.1 applies because $F(x_0, y_0, W)$ has degree m in W. The theorem says that these two polynomials in W have a common factor. Since K is algebraically closed, this common factor vanishes at some w_0, and then we must have $F(x_0, y_0, w_0) = G(x_0, y_0, w_0) = 0$. This proves the first conclusion.

For the second conclusion, suppose that $V(F) \cap V(G)$ contains $mn + 1$ points. Join these points by lines, and pick a point of $\mathbb{P}^2_K$ that is not on any of the lines. We can do so because K, being algebraically closed, is infinite. Applying a projective transformation, we may assume that the point is $[0, 0, 1]$. Write F and G in the form $(*)$. Regarding F and G as polynomials in W, with coefficients in $A = K[X, Y]$, we again form $R(F, G)$, which Lemma 8.4 identifies as a member of $K[X, Y]_{mn}$. For fixed (x_0, y_0), Theorem 8.1 says that $R(F, G)(x_0, y_0) = 0$ if and only if $F(x_0, y_0, W)$ and $G(x_0, y_0, W)$ have a common factor (necessarily a common factor of the form $W - w_0$ because K is algebraically closed), if and only if $F(x_0, y_0, w_0) = G(x_0, y_0, w_0) = 0$ for some w_0. So at each of our $mn + 1$

points, say (x_i, y_i, w_i), we have $R(F, G)(cx_i, cy_i) = 0$ for all scalars c. Since $(x_i, y_i) \neq (0, 0)$, $R(F, G)$ vanishes on the line $y_i X - x_i Y = 0$. Consequently $y_i X - x_i Y$ divides $R(F, G)$ in $K[X, Y]$.

Suppose that (x_i, y_i) is a multiple of (x_j, y_j) with $i \neq j$. Then (x_i, y_i, w_i) and (x_j, y_j, w_j) both satisfy $y_i X - x_i Y = 0$. Since $(0, 0, 1)$ satisfies this also and since $(0, 0, 1)$ is not to be on any of the connecting lines, we obtain a contradiction.

Thus the $mn+1$ factors $y_i X - x_i Y$ are nonassociate primes in $K[X, Y]$ dividing $R(F, G)$. By unique factorization for $K[X, Y]$, their product divides $R(F, G)$. Since $\deg R(F, G) = mn$, we conclude that $R(F, G) = 0$. Then Theorem 8.1 shows that F and G have a nonconstant common factor in $K[X, Y][W] = K[X, Y, W]$. The common factor is homogeneous by Lemma 8.3, and the second conclusion is proved. □

4. Intersection Multiplicity for a Line with a Curve

In this section we begin the topic of "intersection multiplicity" for projective plane curves. The idea is that the number of points in the intersection $V(F) \cap V(G)$ in Bezout's Theorem as formulated in Theorem 8.5 should actually equal mn, not merely be bounded above by mn, if the field is algebraically closed and the points are counted according to their "multiplicities," whatever that might mean.

The prototype is the factorization of a polynomial of degree n in one variable. The polynomial has at most n roots, and it has exactly n if the field is algebraically closed and each root is counted according to its multiplicity. In this case, as we well know, a root z_0 of $f(z)$ has multiplicity k if $(z - z_0)^k$ is the largest power of $z - z_0$ that divides $f(z)$.

Our objective in this section is to develop a notion of intersection multiplicity for the case of a line and a curve at a point; the case of two curves is less intuitive and is postponed to the next section. The main result is to be that the sum of the intersection multiplicities at all points for a line and a projective plane curve equals the degree of the curve, provided that the underlying field is algebraically closed and that the line does not divide the curve. The statement in the previous paragraph about polynomials in one variable will amount to a special case; for this special case the projective line is Y, the projective curve is of the form $W^{d-1}Y - F(X, W)$, where F is homogeneous of degree d and where $f(X) = F(X, 1)$, and the divisibility proviso is that F not be the 0 polynomial, i.e., that $f(z)$ not be identically 0.

Let K be a field, let L be in $K[X, Y, W]_1$, and let F be in $K[X, Y, W]_d$. The notation for intersection multiplicity will be $I(P, L \cap F)$, where $P = (x_0, y_0, w_0)$ is in $V_K(F) \cap V_K(L)$. To make the definition, we introduce affine local coordinates. Choose Φ in $\mathrm{GL}(3, K)$ with $\Phi(x_0, y_0, w_0) = (0, 0, 1)$, and form the corresponding local expressions

$$\begin{aligned} f(X, Y) &= F(\Phi^{-1}(X, Y, 1)) = f_1(X, Y) + \cdots + f_d(X, Y), \\ l(X, Y) &= L(\Phi^{-1}(X, Y, 1)). \end{aligned}$$

Here f_j is the part of f that is homogeneous of degree j. Since $l(0, 0) = 0$, we see that $l(X, Y) = bX - aY$ for some constants a and b not both 0. Then $\varphi(t) = \begin{pmatrix} at \\ bt \end{pmatrix}$, for $t \in K$, is a parametrization of the locus in $\mathbb{A}^2_K$ on which $l(x, y) = 0$. The composition $f(\varphi(t))$ is a polynomial in t with $f(\varphi(0)) = 0$. In fact,

$$\begin{aligned} f(\varphi(t)) &= f_1(at, bt) + f_2(at, bt) + \cdots + f_d(at, bt) \\ &= t f_1(a, b) + t^2 f_2(a, b) + \cdots + t^d f_d(a, b). \end{aligned}$$

There are two possibilities. If $f \circ \varphi$ is not the 0 polynomial, then $f(\varphi(t))$ has a zero of some finite order at $t = 0$, and this order is defined to be the **intersection multiplicity**, or **intersection number**, $I(P, L \cap F)$. If $f \circ \varphi$ is the 0 polynomial, then we say that $I(P, L \cap F) = +\infty$. It will be convenient to define $I(P, L \cap F) = 0$ if P is not in $V_K(L) \cap V_K(F)$. We need to check that $I(P, L \cap F)$ does not depend on the choice of Φ, but we postpone this verification until after we consider two examples.

EXAMPLES OF INTERSECTION MULTIPLICITY.

(1) Example 1 in the previous section showed that relative to a suitable Φ in $\mathrm{GL}(3, K)$, the projective plane curve

$$F(X, Y, W) = X^2Y + XYW + W^3$$

has local expression $f(X, Y)$ about $P = (x_0, y_0, 1)$ given by

$$\begin{aligned} f(X, Y) &= (x_0^2 y_0 + x_0 y_0 + 1) + (x_0^2 Y + 2x_0 y_0 X + x_0 Y + y_0 X) \\ &\quad + (y_0 X^2 + 2x_0 XY + XY) + (X^2 Y) \\ &= f_0 + f_1(X, Y) + f_2(X, Y) + f_3(X, Y). \end{aligned}$$

For a line L, the intersection multiplicity $I(P, L \cap F)$ is 0 unless P lies in $V_K(F)$, i.e., unless $f_0 = x_0^2 y_0 + x_0 y_0 + 1 = 0$. Suppose that the line L is given by

$$L(X, Y, W) = \alpha X + \beta Y + \gamma W,$$

with local expression

$$l(X, Y) = L(X + x_0, Y + y_0, 1) = (\alpha x_0 + \beta y_0 + \gamma) + (\alpha X + \beta Y).$$

Here α and β are not both 0. The intersection multiplicity $I(P, L \cap F)$ is 0 unless P lies also in $V_K(L)$, i.e., unless $\alpha x_0 + \beta y_0 + \gamma = 0$. Thus suppose that P lies in $V_K(L) \cap V_K(F)$. Then we can parametrize the locus for which $l(x, y) = 0$ by $\begin{pmatrix} x \\ y \end{pmatrix} = \varphi(t) = \begin{pmatrix} -\beta t \\ \alpha t \end{pmatrix}$, and we obtain

$$\begin{aligned} f_1(\varphi(t)) &= f_1(-\beta t, \alpha t) = t(x_0^2\alpha - 2x_0 y_0 \beta + x_0\alpha - y_0\beta), \\ f_2(\varphi(t)) &= f_2(-\beta t, \alpha t) = t^2(y_0\beta^2 - 2x_0\alpha\beta + \alpha\beta). \end{aligned}$$

One point lying in $V_K(F)$ is $P = (x_0, y_0, 1) = \left(1, -\frac{1}{2}, 1\right)$, and P lies also in $V_K(L)$ if $\alpha - \frac{1}{2}\beta + \gamma = 0$, i.e., if γ satisfies $\gamma = \frac{1}{2}\beta - \alpha$. Then we have $f_1(\varphi(t)) = t(2\alpha + \frac{3}{2}\beta)$ and $f_2(\varphi(t)) = t^2(-\frac{1}{2}\beta^2 - \alpha\beta)$. Consequently, $I(P, L \cap F)$ is ≥ 1 if and only if $\gamma = \frac{1}{2}\beta - \alpha$. In this case, $I(P, L \cap F)$ is ≥ 2 if and only if $2\alpha + \frac{3}{2}\beta = 0$, i.e., if $\alpha = -\frac{3}{4}\beta$. When both conditions are satisfied, we have $f_2(\varphi(t)) = t^2(-\frac{1}{2}\beta^2 - \alpha\beta) = t^2(\frac{1}{4}\beta^2)$, and this is not the 0 function because under these conditions, $\beta = 0$ would imply that $(\alpha, \beta, \gamma) = (0, 0, 0)$; hence $I(P, L \cap F) = 2$.

(2) Example 2 in the previous section considered the point $P = (x_0, y_0, w_0) = (0, 1, 0)$ for the same F, namely $F(X, Y, W) = X^2Y + XYW + W^3$. This P lies in $V_K(F)$. For a suitable Φ, the earlier computations showed that the local expression for F is

$$f(X, Y) = (Y^2 + XY) + (X^3).$$

The most general line L for which P lies in $V_K(L)$ is $\alpha X + \gamma W = 0$, and the corresponding local expression is

$$l(X, Y) = L(Y, 1, X) = \alpha Y + \gamma X.$$

We use the parametrization $\varphi(t) = (-\alpha t, \gamma t)$ for L and obtain

$$f(\varphi(t)) = t^2(\gamma^2 - \alpha\gamma) + t^3(-\alpha^3).$$

By inspection we see that $I(P, L \cap F) \geq 2$ for all choices of α and γ, and that $I(P, L \cap F) \geq 3$ if and only if $\gamma = 0$ or $\gamma = \alpha$. If $\gamma = 0$ or $\gamma = \alpha$, then α^3 cannot be 0, and thus $I(P, L \cap F) = 3$.

Let us return to the verification that $I(P, L \cap F)$ does not depend on the choice of Φ. Thus suppose that Ψ is another member of $\mathrm{GL}(3, K)$ with $\Psi(x_0, y_0, w_0) = (0, 0, 1)$. Write

$$\Psi \circ \Phi^{-1} = \begin{pmatrix} \alpha & \beta & 0 \\ \gamma & \delta & 0 \\ r & s & 1 \end{pmatrix},$$

form the local expressions

$$\begin{aligned} f'(X,Y) &= F(\Psi^{-1}(X,Y,1)) = f_1'(X,Y) + \cdots + f_d'(X,Y), \\ l'(X,Y) &= L(\Psi^{-1}(X,Y,1)) = b'X - a'Y, \end{aligned}$$

and parametrize the locus in $\mathbb{A}^2_K$ with $l'(x,y) = 0$ by

$$\begin{pmatrix} x \\ y \end{pmatrix} = \varphi'(t) = \begin{pmatrix} a't \\ b't \end{pmatrix}.$$

We need a lemma.

Lemma 8.6. In the above notation, $f(X,Y)$ equals

$$\begin{aligned} &(rX+sY+1)^{d-1} f_1'(\alpha X + \beta Y, \gamma X + \delta Y) \\ &\qquad + (rX+sY+1)^{d-2} f_2'(\alpha X + \beta Y, \gamma X + \delta Y) \\ &\qquad + \cdots + f_d'(\alpha X + \beta Y, \gamma X + \delta Y), \end{aligned}$$

and therefore

$$f_1(X,Y) = f_1'(\alpha X + \beta Y, \gamma X + \delta Y).$$

PROOF. For the first conclusion, let us justify the following computation:

$$\begin{aligned} f(X,Y) &= (F \circ \Psi^{-1})(\Psi \circ \Phi^{-1})(X,Y,1) \\ &= (F \circ \Psi^{-1})(\alpha X + \beta Y, \gamma X + \delta Y, rX + sY + 1) \\ &= (F \circ \Psi^{-1})\Big((rX+sY+1)\big(\tfrac{\alpha X + \beta Y}{rX+sY+1}, \tfrac{\gamma X + \delta Y}{rX+sY+1}, 1\big)\Big) \\ &= (rX+sY+1)^d f'\big(\tfrac{\alpha X + \beta Y}{rX+sY+1}, \tfrac{\gamma X + \delta Y}{rX+sY+1}\big) \\ &= (rX+sY+1)^d (f_1' + \cdots + f_d')\big(\tfrac{\alpha X + \beta Y}{rX+sY+1}, \tfrac{\gamma X + \delta Y}{rX+sY+1}\big) \\ &= (rX+sY+1)^{d-1} f_1'(\alpha X + \beta Y, \gamma X + \delta Y) \\ &\quad + (rX+sY+1)^{d-2} f_2'(\alpha X + \beta Y, \gamma X + \delta Y) \\ &\quad + \cdots + f_d'(\alpha X + \beta Y, \gamma X + \delta Y). \end{aligned}$$

In fact, the first three lines are valid if we make the computation in the field of fractions $K(X,Y)$, the fourth line uses the homogeneity of F and a substitution homomorphism that evaluates members of $K[X,Y,W]$ at points of $K(X,Y,W)$, and the remaining lines use the homogeneity of $f_1', \ldots, f_d'$ and a substitution homomorphism that evaluates their arguments at points of $K(X,Y)$.

This proves the first conclusion. To derive the second conclusion from it, we expand each of the coefficients on the right side and group terms of the same degree of homogeneity under $(X, Y) \mapsto (\lambda X, \lambda Y)$. The only term whose degree of homogeneity is 1 is $f'_1(\alpha X+\beta Y, \gamma X+\delta Y)$ with a coefficient 1 coming from the expansion of $(rX+sY+1)^{d-1}$; all other terms have higher degree of homogeneity. When $f(X, Y)$ on the left side is expanded as a sum of homogeneous polynomials, the term of degree 1 is $f_1(X, Y)$. The second conclusion follows. □

Continuing with the verification that $I(P, L \cap F)$ does not depend on the choice of Φ, we apply Lemma 8.6 to L in place of F, and we obtain

$$l(X, Y) = l'(\alpha X + \beta Y, \gamma X + \delta Y).$$

Since $l(X, Y) = bX - aY$ and $l'(X, Y) = b'X - a'Y$, this equation shows that

$$b = b'\alpha - a'\gamma \qquad \text{and} \qquad -a = b'\beta - a'\delta.$$

Putting $\Delta = \alpha\delta - \beta\gamma$, we solve for a' and b' and obtain

$$\alpha a + \beta b = \Delta a' \qquad \text{and} \qquad \gamma a + \delta b = \Delta b'.$$

When $x = at$ and $y = bt$, we thus have

$$\alpha x + \beta y = \alpha at + \beta bt = t\Delta a' \qquad \text{and} \qquad \gamma x + \delta y = \gamma at + \delta bt = t\Delta b'.$$

Substituting these formulas into the first conclusion of Lemma 8.6 and using the homogeneity of each f'_j gives

$$\begin{aligned} f(\varphi(t)) = {} & (art + bst + 1)^{d-1} t\Delta f'_1(a', b') \\ & + (art + bst + 1)^{d-2} t^2 \Delta^2 f'_2(a', b') + \cdots + t^d \Delta^d f'_d(a', b'). \end{aligned}$$

If j is the smallest index for which $f'_j(a', b') \neq 0$, then the lowest power of t remaining on the right side after expansion of the coefficients is t^j, and its coefficient is $\Delta^j f'_j(a', b')$. Thus we can conclude that the lowest power of t with nonzero coefficient on the left side is t^j, and its coefficient $f_j(a, b)$ must equal $\Delta^j f'_j(a', b')$. The equality of the lowest power of t remaining on each side shows that $I(P, L \cap F)$ is the same when computed from f as when computed from f', and we obtain as a bonus the formula $f_j(a, b) = \Delta^j f'_j(a', b')$ if t^j is that power. This completes the verification that $I(P, L \cap F)$ does not depend on the choice of Φ.

Now we come back to the circle of ideas around Bezout's Theorem. The first task is to clarify the meaning of infinite intersection multiplicity.

Proposition 8.7. Over the field K if a projective line L and a projective plane curve F meet at a point P in $\mathbb{P}^2_K$, then $I(P, L \cap F) = +\infty$ if and only if L divides F.

PROOF. If L divides F, then in the above notation the local expression $l(X, Y)$ divides $f(X, Y)$. Since $l(\varphi(t))$ is the 0 polynomial, so is $f(\varphi(t))$.

Conversely suppose that $f(\varphi(t))$ is the 0 polynomial, so that $f_r(a, b) = 0$ for all r with $1 \leq r \leq d = \deg F$. Without loss of generality, suppose $b \neq 0$. The equality

$$\begin{aligned} 0 = f_r(a, b) &= c_0 a^r + c_1 a^{r-1} b + \cdots + c_r b^r \\ &= b^r \big(c_0 (ab^{-1})^r + c_1 (ab^{-1})^{r-1} + \cdots + c_r\big) \end{aligned}$$

says that $Z - ab^{-1}$ is a factor of $b^r(c_0 Z^r + c_1 Z^{r-1} + \cdots + c_r)$. If we write

$$b^r(c_0 Z^r + c_1 Z^{r-1} + \cdots + c_r) = (Z - ab^{-1})u(Z)$$

and take $Z = XY^{-1}$, then

$$\begin{aligned} b^r f_r(X, Y) &= b^r Y^r \big(c_0 (XY^{-1})^r + c_1 (XY^{-1})^{r-1} + \cdots + c_r\big) \\ &= Y^r (XY^{-1} - ab^{-1}) u(XY^{-1}) = b^{-1} l(X, Y) \big(Y^{r-1} u(XY^{-1})\big). \end{aligned}$$

Hence $l(X, Y)$ divides $f_r(X, Y)$ for all r. It follows that $l(X, Y)$ divides $f(X, Y)$ and then that L divides F. □

The full-strength version of Bezout's Theorem says that two projective plane curves F and G of degrees m and n meet in at most mn points even when multiplicities are counted, and that the number is equal to mn if K is algebraically closed and multiplicities are counted. This theorem will be proved in Section 6. For the time being, we shall limit ourselves to the special case of the full-strength theorem in which one of the curves is a line.

Theorem 8.8 (Bezout's Theorem). Let K be an algebraically closed field. If F is a projective plane curve over K of degree d and if L is a projective line such that L does not divide F, then $\sum_P I(P, L \cap F) = d$.

PROOF. First we show that

$$\sum_P I(P, L \cap F) < +\infty. \tag{$*$}$$

Since L is assumed not to divide F, Proposition 8.7 shows that $I(P, L \cap F)$ is finite at every point of $V_K(L) \cap V_K(F)$. Thus $\sum_P I(P, L \cap F)$ is finite if there are only finitely many points in $V_K(L) \cap V_K(F)$. Bezout's Theorem in the form of Theorem 8.5 shows that either $V_K(L) \cap V_K(F)$ is finite or else L and F have as a common factor some homogeneous polynomial of positive degree. Since L has degree 1, L is prime, and thus L and F can have a common factor of positive degree only if L divides F. We are assuming the contrary, and therefore $V_K(L) \cap V_K(F)$ is finite. This proves $(*)$.

Possibly by applying a projective transformation, we may assume[7] that the given line L is the line at infinity W. Then the points P_j with $I(P_j, W \cap F) > 0$ are of the form $[x_j, y_j, 0]$. Taking into account that the algebraically closed field K is necessarily infinite, we can apply a second projective transformation, one that translates the Y variable, and assume that no y_j is 0. Then we can write $P_j = [r_j, 1, 0]$ with r_j in K. Let us see that

$$H(X) = F(X, 1, 0) \quad \text{is a nonzero polynomial of degree exactly } d. \qquad (**)$$

In fact, $F(X, Y, W)$ is homogeneous of degree d, and we have arranged that $[1, 0, 0]$, which certainly lies in $V_K(W)$, is not in $V_K(F)$. Consequently the X^d term in $F(X, Y, W)$ has nonzero coefficient, and $(**)$ follows.

Next let us prove that

$$I\big((r, 1, 0), W \cap F\big) = \text{multiplicity of } r \text{ as a root of } H(X) = F(X, 1, 0). \qquad (\dagger)$$

Then it will follow that $\sum_P I(P, W \cap F)$ equals the number of roots of $H(X) = F(X, 1, 0)$, each counted as many times as its multiplicity. In view of $(**)$ and the fact that K is algebraically closed, we will then have proved that $\sum_P I(P, W \cap F) = d$, as required.

To prove $(\dagger)$, we introduce affine local coordinates about $(r, 1, 0)$, using $\Phi^{-1} = \begin{pmatrix} 1 & 0 & r \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, so that $\Phi(r, 1, 0) = (0, 0, 1)$. The local versions f of F and l of W relative to this Φ are

$$\begin{aligned} f(X, Y) &= F(\Phi^{-1}(X, Y, 1)) = F(X + r, 1, Y), \\ l(X, Y) &= W(\Phi^{-1}(X, Y, 1)) = Y. \end{aligned}$$

Hence $l(X, Y)$ is of the form $bX - aY$ with $a = -1$ and $b = 0$. If we parametrize l by $\varphi(t) = (at, bt) = (-t, 0)$, then

$$f(\varphi(t)) = f(-t, 0) = F(-t + r, 1, 0).$$

[7] If P and P' are distinct points in $\mathbb{P}^2_K$, then there exists a projective transformation carrying P to $[1, 0, 0]$ and P' to $[0, 1, 0]$. This transformation carries the unique line through P and P' to the line at infinity.

The order of vanishing of $f(\varphi(t))$ at $t = 0$, which is $I\big([r, 1, 0], W \cap F\big)$, thus equals the order of the zero of $F(-t + r, 1, 0)$ at $t = 0$, which equals the multiplicity of r as a root of $H(X) = F(X, 1, 0)$. This proves (†), and the theorem follows. □

5. Intersection Multiplicity for Two Curves

In this section we continue the topic of "intersection multiplicity" begun in Section 4. That section dealt with intersection multiplicity for the special case of a projective line and a projective plane curve, and the present section deals with the general case of two projective plane curves. The next section will use the general notion to address Bezout's Theorem in full generality. In this section and the next we shall make occasional use of material from Chapter VII, especially Lemma 7.21 and the results in Section VII.1.

It is worth reviewing qualitatively what happened in Section 4. What we did was refer the given line and curve to affine space, parametrize the line in a natural way, and substitute the parametrization into the formula for the curve to obtain a scalar-valued function of one variable. The order of vanishing of the resulting scalar-valued function of one variable was defined to be the intersection multiplicity. The classical approach[8] for handling two curves proceeds by trying to generalize this construction, in effect parametrizing one curve and substituting into the other. The fact that there need be no natural parametrization of either of the curves leads to a number of complications, and ultimately the argument involves a complicated ring of power series.

We shall follow a somewhat more modern approach[9] based on localizations.[10] The definition is not particularly intuitive, and it is necessary to study some examples to see its virtues. We give the definition, show that the definition is consistent with the definition in the special case of Section 4, check that the definition makes sense in general, state some properties that are useful in making computations, work out an example, and then verify the properties. Thus let F and G be homogeneous polynomials in (X, Y, W) of respective degrees m and n, and let $P = [x_0, y_0, w_0]$ be a point of the projective plane $\mathbb{P}^2_K$ over a field K. We refer matters back to affine space in the usual way by letting Φ be any member of $\mathrm{GL}(3, K)$ such that $\Phi(x_0, y_0, w_0) = (0, 0, 1)$. The local expressions from Φ

[8] An account appears in Walker, Chapter IV.

[9] See Fulton, Chapter 3, for the present section and Fulton, Chapter 5, for the next section.

[10] For a still more modern and more general approach, see Serre's *Algèbre Locale*. Serre's opening sentence summarizes matters by saying, "Intersection multiplicities in algebraic geometry are equal to certain 'Euler–Poincaré characteristics' formed by means of the Tor functors of Cartan–Eilenberg."

about $(0, 0)$ corresponding to F and G are the polynomials f and g with

$$f(X, Y) = F(\Phi^{-1}(X, Y, 1)),$$
$$g(X, Y) = G(\Phi^{-1}(X, Y, 1)).$$

These polynomials break into homogeneous parts as

$$f(X, Y) = f_0 + f_1(X, Y) + \cdots + f_m(X, Y),$$
$$g(X, Y) = g_0 + g_1(X, Y) + \cdots + g_n(X, Y),$$

with f_j and g_j homogeneous of degree j in the pair (X, Y). We assume that P lies on the locus $V_K(F) \cap V_K(G)$ of common zeros of F and G, and the condition for this to happen is that $f_0 = g_0 = 0$. The **order of vanishing** $m_P(F)$ of F at P is the first j for which f_j is not the zero polynomial; we saw as a consequence of Lemma 8.6 that this quantity is well defined independently of the choice of Φ.

The **intersection multiplicity** $I(P, F \cap G)$ of F and G at P can be defined in either of two equivalent ways. The equivalence of the two definitions will be used repeatedly in the discussion and follows from the fact that localization commutes with passage to the quotient by an ideal, a fact that was proved as Lemma 7.21. One definition is

$$I(P, F \cap G) = \dim_K \big(\big(K[X, Y]/(f, g)\big)_{(0,0)}\big),$$

where $\big(K[X, Y]/(f, g)\big)_{(0,0)}$ is the localization at $(0, 0)$ of the K algebra $K[X, Y]/(f, g)$. That is, we form the quotient ring of $K[X, Y]$ by the ideal generated over K by f and g, localize with respect to the maximal ideal of all members of the quotient vanishing at $(0, 0)$, and compute the dimension of this localization over K. The other definition is

$$I(P, F \cap G) = \dim_K \big(S^{-1}K[X, Y]\big/S^{-1}(f, g)\big),$$

where S is the multiplicative system in $K[X, Y]$ consisting of the complement of the maximal ideal (X, Y), i.e., consisting of all polynomials that are nonvanishing at $(0, 0)$. In either case all elements of the ring being localized have interpretations as functions, and the multiplicative system consists of all the functions that are nonzero at a certain point. Nevertheless, the matter is a little subtle because some members of the multiplicative system in the first case may be zero divisors. Here is a lower-dimensional example of that phenomenon that can also serve as a guiding example for Theorem 8.12 below.

EXAMPLE OF GEOMETRIC LOCALIZATION. $R = \big(K[X]/((X^2(X-1)^2))\big)_{(0)}$, with the subscript indicating localization at 0. Before passage to the localization, the quotient $Q = K[X]/((X^2(X-1)^2))$ has dimension 4, with a basis consisting of the cosets of $1, X, X^2, X^3$. The multiplicative system S for localization at 0 consists of all members of the quotient that are nonzero at 0. The localization as a set consists of equivalence classes of pairs (r, s) with r in Q and s in S, two pairs (r, s) and (r', s') being equivalent if $t(rs' - r's) = 0$ for some t in S. Localization is a ring homomorphism, and we therefore consider the pairs (r, s) in the class of the additive identity. These have $t(r1 - 0s) = 0$ for some t. Then t and r have representatives $t(X)$ and $r(X)$ in $K[X]$ such that $t(X)r(X) = p(X)X^2(X-1)^2$ for some $p(X)$. Furthermore, $t(0) \neq 0$. Then X^2 must divide $r(X)$, and this condition is also sufficient for the choice $t(X) = (X-1)^2$. Thus the members $X^2q(X)$ of $K[X]$ give 0 in the localization, and the localization is isomorphic to the 2-dimensional algebra $K[X]/(X^2)$.

Proposition 8.9 below will show that $I(P, F \cap G)$ is independent of the function Φ used to introduce affine local coordinates. Assuming this independence, we begin with an example that shows that the definition is consistent with the definition in Section 4.

EXAMPLE 1 OF INTERSECTION MULTIPLICITY. Case of a line L and a curve F homogeneous of degree d. Assuming that P lies in $V_K(L) \cap V_K(F)$, we introduce affine local coordinates by means of a member Φ of $\mathrm{GL}(3, K)$ that carries a representative of P to $(0, 0, 1)$, and we let $l(X, Y)$ and $f(X, Y)$ be the corresponding local expressions for L and F. Let $f = f_1 + \cdots + f_d$ be the decomposition of f into its homogeneous parts. Since the intersection multiplicity is being assumed to be independent of the choice of Φ and since for any second point on a line through $(0, 0, 1)$, there exists a Φ that fixes $(0, 0, 1)$ and carries that second point to $(1, 0, 1)$, we may assume that $l(X, Y) = Y$. We introduce the parametrization $(x, y) = \varphi(t) = (t, 0)$ for the line $l(X, Y)$ and substitute into $f(X, Y)$, obtaining $f(\varphi(t)) = f_1(t, 0) + \cdots + f_d(t, 0)$. In the definition of Section 4, the intersection multiplicity is the least r such that $f_r(t, 0)$ is not identically 0, or else it is $+\infty$ if $f(\varphi(t))$ is identically 0. With the new definition we observe from the definition of r that f is of the form

$$f(X, Y) = (c_rX^r + \cdots + c_dX^d) + Yg(X, Y) = c_rX^r(1 + Xh(X)) + Yg(X, Y)$$

with $c_r \neq 0$, $g(X, Y) \in K[X, Y]$, and $h(X) \in K[X]$. The ideal in $K[X, Y]$ generated by Y and f is the same as the ideal generated by Y and $X^r(1 + Xh(X))$. Hence

$$K[X, Y]/(Y, f) \cong K[X, Y]/(Y, X^r(1 + Xh)) \cong K[X]/(X^r(1 + Xh)).$$

The polynomial $1 + Xh(X)$ takes a nonzero value at 0 and hence is a member of the multiplicative system that we use to form the localization. Thus

$$\big(K[X,Y]/(Y,f)\big)_{(0,0)} \cong \big(K[X]/(X^r(1+Xh))\big)_{(0)} \cong \big(K[X]/(X^r)\big)_{(0)}.$$

The dimension of the right side is r, and thus the new definition of intersection multiplicity matches the old one.

Proposition 8.9. The intersection multiplicity of two projective plane curves F and G at P is well defined independently of the member of Φ that moves a representative of P to $(0, 0, 1)$.

PROOF. It is enough to take $P = [0, 0, 1]$ and to compare the effect of passing to affine local coordinates determined by the identity with the effect of passing to the coordinates determined by a general element Φ of $\mathrm{GL}(3, K)$ of the form $\Phi = \begin{pmatrix} \alpha & \beta & 0 \\ \gamma & \delta & 0 \\ r & s & 1 \end{pmatrix}$. Let $\deg F = m$ and $\deg G = n$. If $f(X,Y) = F(X,Y,1)$ and $\widetilde{f}(X,Y) = F(\Phi^{-1}(X,Y,1))$, then the computation in the proof of Lemma 8.6 shows that

$$f(X,Y) = (1 + rX + sY)^m\, \widetilde{f}\big(\tfrac{\alpha X+\beta Y}{1+rX+sY}, \tfrac{\gamma X+\delta Y}{1+rX+sY}\big). \qquad (*)$$

Similarly if $g(X,Y) = G(X,Y,1)$ and $\widetilde{g}(X,Y) = G(\Phi^{-1}(X,Y,1))$, then

$$g(X,Y) = (1 + rX + sY)^n\, \widetilde{g}\big(\tfrac{\alpha X+\beta Y}{1+rX+sY}, \tfrac{\gamma X+\delta Y}{1+rX+sY}\big).$$

Let

$$X' = \tfrac{\alpha X+\beta Y}{1+rX+sY}, \quad Y' = \tfrac{\gamma X+\delta Y}{1+rX+sY}, \quad \text{and} \quad \Phi^{-1} = \begin{pmatrix} \alpha' & \beta' & 0 \\ \gamma' & \delta' & 0 \\ r' & s' & 1 \end{pmatrix}.$$

It is purely a formal matter that the mapping T defined by $(Th)(X,Y) = h(X',Y')$ is a field isomorphism of $K(X,Y)$ onto $K(X',Y')$. It sends $K[X,Y]$ onto $K[X',Y']$ and sends $\big(K[X,Y]\big)_{(0,0)}$ onto $\big(K[X',Y']\big)_{(0,0)}$. Referring to the formulas for X' and Y', we see that the image of $K[X,Y]$ is contained in the localization $\big(K[X,Y]\big)_{(0,0)}$; by the universal mapping property of localizations, the image of $\big(K[X,Y]\big)_{(0,0)}$ is contained in $\big(K[X,Y]\big)_{(0,0)}$. Comparing these two conclusions, we see that $\big(K[X',Y']\big)_{(0,0)} \subseteq \big(K[X,Y]\big)_{(0,0)}$.

Meanwhile, we can solve the equations defining X' and Y' for X and Y. If we compare the results with the formula for Φ^{-1}, we find that

$$X = \tfrac{\alpha' X'+\beta' Y'}{1+r'X'+s'Y'} \quad \text{and} \quad Y = \tfrac{\gamma' X'+\delta' Y'}{1+r'X'+s'Y'}.$$

Thus the situation is symmetric, and we have $\big(K[X,Y]\big)_{(0,0)} \subseteq \big(K[X',Y']\big)_{(0,0)}$. Consequently the mapping

$$(Th)(X,Y) = h\big(\tfrac{\alpha X+\beta Y}{1+rX+sY}, \tfrac{\gamma X+\delta Y}{1+rX+sY}\big)$$

is an algebra automorphism of $\big(K[X,Y]\big)_{(0,0)}$.

To prove the proposition, recall that localization commutes with passage to the quotient by an ideal. In view of $(*)$, it is therefore enough to show that

$$\begin{aligned}&\dim_K\,\big(\big(K[X,Y]\big)_{(0,0)}\big/(f,g)\big)\\ &\quad\overset{?}{=} \dim_K\,\big(\big(K[X,Y]\big)_{(0,0)}\big/((1+rX+sY)^m Tf, (1+rX+sY)^n Tg)\big). \qquad (**)\end{aligned}$$

The factor $(1+rX+sY)$ is a unit in $\big(K[X,Y]\big)_{(0,0)}$, and we can simplify the quotient algebra on the right side of $(**)$ to

$$\big(K[X,Y]\big)_{(0,0)}\big/(Tf, Tg).$$

In turn, this algebra is K isomorphic to $\big(K[X,Y]\big)_{(0,0)}\big/(f,g)$ because T is an automorphism of $\big(K[X,Y]\big)_{(0,0)}$. The dimensional equality in $(**)$ follows. □

Let us extend the definition of intersection multiplicity to include the case that the point of interest does not lie in the locus of common zeros. We define $I(P, F\cap G) = 0$ if P is not in $V_K(F)\cap V_K(G)$. Assume now that K is algebraically closed. Below we compute a fairly typical example of intersection multiplicity. To do so, we shall make use of certain properties of $I(P, F\cap G)$ that we list in Theorem 8.10 below. In fact, there is an algorithm for computing $I(P, F\cap G)$ using only these properties,[11] but we shall not give it.

Before stating the properties, we need to make some definitions. Recall from earlier in the section that the order of vanishing $m_P(G)$ of G at P is computed using a suitable Φ in GL$(3,K)$ to refer G to affine local coordinates about P, defining $g(X,Y) = G(\Phi^{-1}(X,Y,1))$, expanding $g(X,Y)$ as a sum of homogeneous terms $g(X,Y) = g_0 + g_1(X,Y) + \cdots + g_n(X,Y)$, and defining $m_P(G)$ to be the least j such that g_j is not the 0 polynomial. The homogeneous polynomial $g_j(X,Y)$ is X^j times a polynomial in the one variable YX^{-1}, and the fact that K is algebraically closed implies that g_j has a factorization of the form

$$g_j(X,Y) = c\prod_i (\alpha_i X + \beta_i Y)^{m_i}$$

[11] Fulton, p. 76.

with c in K. Here $j = \sum_i m_i$, and the pairs (α_i, β_i) correspond to distinct members of $\mathbb{P}^1_K$ that are uniquely determined up to indexing if $c \neq 0$. Let $l_i(X, Y) = \alpha_i X + \beta_i Y$, and let L_i be the corresponding projective line. We refer to all the lines L_i as the **tangent lines** to G at P, and we say that m_i is the **multiplicity** of L_i. The geometry of the situation is indicated in Problem 12 at the end of the chapter.

Theorem 8.10. Let K be an algebraically closed field, let P be in $\mathbb{P}^2_K$, and let F and G be projective plane curves over K. Then the intersection multiplicity $I(P, F \cap G)$ has the following properties:

(a) $I(P, F \cap G) = I(P, G \cap F)$,

(b) $I(P, F \cap G) = I(P, F \cap (G + HF))$ for any projective plane curve H with $\deg HF = \deg G$ such that $G + HF \neq 0$,

(c) $I(P, F \cap G) > 0$ if and only if P lies in $V_K(F) \cap V_K(G)$,

(d) $I(P, F \cap G) \leq I(P, AF \cap BG)$ for any projective plane curves A and B, with equality if A and B are nonvanishing at P,

(e) $I(P, F \cap G)$ is finite if and only if F and G have no common factor of degree ≥ 1 having P on its zero locus,

(f) $I(P, F \cap GH) = I(P, F \cap G) + I(P, F \cap H)$ and consequently if $F = \prod_i F_i^{r_i}$ and $G = \prod_j G_j^{s_j}$, then $I(P, F \cap G) = \sum_{i,j} r_i s_j I(P, F_i \cap G_j)$,

(g) $I(P, F \cap G) \geq m_P(F) m_P(G)$, with equality if F and G have no tangent lines in common at P.

REMARKS. Properties (a) and (b) are evident. Properties (c) and (d) are conversational and will be proved in these remarks. Properties (e), (f), and (g) require proofs, and we give those proofs after computing an example. For (c), if P lies in $V_K(F) \cap V_K(G)$, then the local expressions $f(X, Y)$ and $g(X, Y)$ vanish at 0, and so does every member of the ideal (f, g); therefore (f, g) is a proper ideal in $\big(K[X, Y]\big)_{(0,0)}$, and the dimension of the quotient is positive. Conversely if P is not in $V_K(F)$, say, then $f(X, Y)$ lies in the multiplicative system S of nonvanishing polynomials at $(0, 0)$, and $S^{-1}(f, g) = (1)$; hence $S^{-1}K[X, Y]/S^{-1}(f, g) = 0$, and $I(P, F \cap G) = 0$. For (d), $S^{-1}(af, bg) \subseteq S^{-1}(f, g)$ with equality if a and b are nonvanishing at $(0, 0)$, and hence $S^{-1}K[X, Y]/S^{-1}(f, g)$ is a homomorphic image of $S^{-1}K[X, Y]/S^{-1}(af, bg)$ and is a one-one homomorphic image if a and b are nonvanishing at $(0, 0)$.

EXAMPLE 2 OF INTERSECTION MULTIPLICITY. Let $K = \mathbb{C}$, and let the two projective curves be the homogeneous versions of $Y^2 = X^3$ and $Y^2 = X^5$. In other words, let

$$F(X, Y, W) = Y^2W - X^3 \qquad \text{and} \qquad G(X, Y, W) = Y^2W^3 - X^5.$$

We compute $I(P, F \cap G)$ for all points P in $V_K(F) \cap V_K(G)$. In the affine plane the intersections (x, y) may be found by substituting the one equation into the other (or, with more effort in this case, by using the resultant). We obtain $x^5 - x^3 = 0$. This gives $x^3(x^2 - 1) = 0$. The factor $x^2 - 1$ has two distinct roots, and each gives two distinct y's. Thus we obtain the five affine solutions $(+1, \pm 1)$, $(-1, \pm i)$, $(0, 0)$. The fact that the first four occurred routinely with multiplicity 1 translates into intersection multiplicity 1 for each: In fact, (b) shows that $I(P, F \cap G) = I(P, F \cap (W^2F - G))$, and $W^2F - G$ restricts at $(X, Y, 1)$ to $X^5 - X^3 = X^3(X^2 - 1)$. At each of the points $(+1, \pm 1)$, $X^5 - X^3$ when viewed as equal to 0 has a vertical tangent $X - 1$ of multiplicity 1, while $Y^2 - X^3$ has a tangent that is not vertical. A similar argument applies at each of the points $(-1, \pm i)$. By (g), the intersection multiplicity is 1 at each of the four points $(+1, \pm 1)$ and $(-1, \pm i)$.

Next let us consider $(0, 0)$. The order of $X^5 - X^3$ is 3, and the homogeneous term of degree 3, namely $-X^3$, factors as the cube of a linear factor that gives the vertical line X. Meanwhile, $Y^2 - X^3$ has order 2 at $(0, 0)$, and Y^2 factors as the square of a linear factor that gives the horizontal line Y. The two curves have no tangents in common. Hence equality holds in (g), and the intersection multiplicity is 6 at $(0, 0)$.

Finally let us check points (x, y, w) on the line at infinity, i.e., those with $w = 0$. Putting $w = 0$ in the formula $F = G = 0$ shows that $x = 0$. Thus the only point of $V_K(F) \cap V_K(G)$ on the line at infinity is $P = [x_0, y_0, w_0] = [0, 1, 0]$. The local versions of F and G may be given in the variables X and W by restricting (X, Y, W) to $(X, 1, W)$ and considering the polynomials about $(x, w) = (0, 0)$. As above, (b) gives $I(P, F \cap G) = I(P, F \cap (W^2F - G))$, but $F = Y^2W - X^3$ restricts to $W - X^3$ and $W^2F - G = -W^2X^3 + X^5$ remains unchanged upon restriction. The respective lowest-order terms, in factored form, are W and $-X^3(X + W)(X - W)$. None of the factors of the first polynomial matches a factor of the second polynomial, and (g) says that the intersection multiplicity is $1 \cdot 5 = 5$.

The upshot is that we get multiplicity 6 from $(0, 0)$, multiplicity 1 apiece from four other points in the affine plane, and multiplicity 5 from $P = [0, 1, 0]$. The total is 15, the product of the degrees of the given curves, as it must be if we are to have any chance of obtaining the desired generalization of Bezout's Theorem.

To get at Theorem 8.10, we make use of a structure theorem about ideals I in $K[X_1, \dots, X_n]$ for which $V(I)$ is a finite set. To prove the structure theorem, which appears as Theorem 8.12 below, we first prove a lemma about the radical $\sqrt{I}$ of an ideal I, a notion defined in Section VII.1.

Lemma 8.11. If R is a commutative Noetherian ring and I is an ideal in R, then $(\sqrt{I}\,)^m \subseteq I$ for some integer $m \geq 1$.

PROOF. Since R is Noetherian, the ideal $\sqrt{I}$ is finitely generated. Let $\{a_1, \dots, a_n\}$ be a set of generators for it. By definition of radical, choose integers $k_1, \dots, k_n$ such that $a_j^{k_j}$ is in I for $1 \le j \le n$, and put $m = \sum_{j=1}^n k_j$. The most general element of $\sqrt{I}$ is of the form $\sum_{j=1}^n r_j a_j$ with all r_j in R. The m^{th} power of this element is a sum of terms of the form $r a_1^{l_1} \cdots a_n^{l_n}$ with $\sum_{j=1}^n l_j = m$. In view of the definition of m, we must have $l_j \ge k_j$ for some j. Then the factor $a_j^{l_j}$ is in I, and hence the whole term $r a_1^{l_1} \cdots a_n^{l_n}$ is in I. □

Theorem 8.12. Let K be an algebraically closed field, and let I be an ideal in the polynomial ring $K[X_1, \dots, X_n]$ whose locus of common zeros in K^n is a finite set $\{P_1, \dots, P_k\}$. Then $K[X_1, \dots, X_n]/I$ is isomorphic as a ring to the product of its localizations at the points P_j:

$$K[X_1, \dots, X_n]/I \cong \prod_{j=1}^k \big(K[X_1, \dots, X_n]/I\big)_{(P_j)}.$$

Consequently

$$\dim_K(K[X_1, \dots, X_n]/I) = \sum_{j=1}^k \dim_K \big(K[X_1, \dots, X_n]/I\big)_{(P_j)}.$$

REMARKS. The one-variable case is a guide: The ideal I is principal, and we can write $K[X]/I$ as $K[X]/(\prod_{j=1}^k (X - c_j)^{m_j})$. The points P_j of the theorem are the members c_j of K, and the same argument as for the first example of the section shows that $\big(K[X]/(\prod_j (X-c_j)^{m_j})\big)_{(c_j)} \cong K[X]/(X-c_j)^{m_j}$. The isomorphism of the theorem therefore reduces to an instance of the Chinese Remainder Theorem.

PROOF. Let $\varphi_j : K[X_1, \dots, X_n]/I \to \big(K[X_1, \dots, X_n]/I\big)_{(P_j)}$ be the canonical homomorphism, and let $\varphi = (\varphi_1, \dots, \varphi_k)$. The mapping φ is a ring homomorphism into $\prod_{j=1}^k \big(K[X_1, \dots, X_n]/I\big)_{(P_j)}$, and we shall prove that φ is one-one onto. Doing so requires some preparation.

Let I_j be the maximal ideal of all polynomials vanishing at P_j. The Nullstellensatz (Theorem 7.1) shows that $\sqrt{I}$ consists of all $f \in K[X, Y]$ such that f vanishes at each P_i, i.e., that $\sqrt{I} = \bigcap_{j=1}^k I_j$. Lemma 8.11 shows that $(\sqrt{I}\,)^m \subseteq I$ for some m, and thus $\big(\bigcap_{j=1}^k I_j\big)^m \subseteq I$. For $i \ne j$, $I_i^m + I_j^m$ is an ideal whose locus of common zeros is empty, and the Nullstellensatz shows that $I_i^m + I_j^m = K[X_1, \dots, X_n]$. The Chinese Remainder Theorem (Theorem 8.27 of *Basic Algebra*) therefore applies and shows that the intersection $\bigcap_{j=1}^k I_j^m$ and the product $\prod_{j=1}^k I_j^m$ coincide. Similarly $I_i + I_j = K[X_1, \dots, X_n]$, and hence $\bigcap_{j=1}^k I_j = \prod_{j=1}^k I_j$. Putting these facts together, we conclude that

$$\bigcap_{j=1}^k I_j^m = \prod_{j=1}^k I_j^m = \big(\prod_{j=1}^k I_j\big)^m = \big(\bigcap_{j=1}^k I_j\big)^m \subseteq I. \qquad (*)$$

Let us now denote members of $K[X_1, \dots, X_n]$ by uppercase letters and their cosets modulo I by the corresponding lowercase letters. Let us observe for $1 \le i \le k$ that there exists $F_i \in K[X_1, \dots, X_n]$ with $F_i(P_j) = \delta_{ij}$. In fact, we start from the special case that if $P \ne Q$, then there exists F with $F(P) = 1$ and $F(Q) = 0$. For the special case, P and Q differ in some coordinate; say that $x_l(P) \ne x_l(Q)$. Then the polynomial

$$F(X_1, \dots, X_n) = (X_l - x_l(Q))(x_l(P) - x_l(Q))^{-1}$$

has the required properties. To construct F_1 with $F_1(P_j) = \delta_{1j}$, choose G_j with $G_j(P_1) = 1$ and $G_j(P_j) = 0$. Then $F_1 = \prod_{i \ne 1} G_i$ has $F_1(P_1) = 1$ and $F_1(P_j) = 0$ for $j \ne 1$. The polynomials $F_2, \dots, F_k$ are constructed similarly.

With m as in the second paragraph of the proof, fix j and define $E_i = 1 - (1 - F_i^m)^m$. This is divisible by F_i^m and hence lies in I_j^m if $i \ne j$. In addition, $1 - F_j^m$ lies in I_j, and hence $1 - E_j = (1 - F_j^m)^m$ is in I_j^m. Therefore $1 - \sum_{i=1}^k E_i = (1 - E_j) - \sum_{i \ne j} E_i$ lies in I_j^m. Since the left side is independent of j, $1 - \sum_{i=1}^k E_i$ lies in $\bigcap_{j=1}^k I_j^m$, and we conclude from $(*)$ that

$$1 - \sum_{i=1}^k E_i \qquad \text{lies in } I. \qquad (**)$$

We just saw that E_i lies in $\bigcap_{j \ne i} I_j^m$. Hence if $i \ne j$, then $E_i E_j$ lies in $\bigcap_{l=1}^k I_l^m \subseteq I$. Passing to cosets modulo I, we find from this fact and from $(**)$ that

$$e_i e_j = 0 \ \text{ for } i \ne j, \qquad \text{and that} \qquad \sum_{i=1}^k e_i = 1. \qquad (\dagger)$$

Multiplying the second equation by e_j and substituting from the first equation, we obtain

$$e_i^2 = e_i \qquad \text{for all } i. \qquad (\dagger\dagger)$$

Using $(\dagger)$ and $(\dagger\dagger)$, let us prove for each i that

$$\text{to each } G \in K[X_1, \dots, X_n] \text{ with } G(P_i) \ne 0 \text{ corresponds a polynomial } H \text{ with } hg = e_i. \qquad (\ddagger)$$

In fact, we may assume that $G(P_i) = 1$. Let Q be the member of I_i given by $Q = 1 - G$. The element $Q^m E_i$ is in I_i^m because Q is in I_i, and it is in I_j^m for $j \neq i$ because E_i is in I_j^m for $j \neq i$. Thus $Q^m E_i$ is in $\bigcap_{j=1}^k I_j^m \subseteq I$, and $q^m e_i = 0$. Consequently

$$g(e_i + qe_i + \cdots + q^{m-1}e_i) = (1-q)e_i(1+q+\cdots+q^{m-1}) = e_i(1-q^m) = e_i,$$

and $H = E_i(1 + Q + \cdots + Q^{m-1})$ is a polynomial as in (‡).

Now we can prove that φ is one-one. If f is a member of $K[X_1, \dots, X_n]/I$ such that $\varphi(f) = 0$, then $\varphi_i(f) = 0$ for all i. This means that there exists a member g_i of the multiplicative system for localization at P_i such that $g_i f = 0$. Any corresponding polynomial G_i has $G_i(P_i) \neq 0$. By (‡), there exists h_i with $h_i g_i = e_i$. Then (†) gives $f = \sum_{i=1}^k e_i f = \sum_{i=1}^k h_i g_i f = 0$. Thus φ is one-one.

For the proof that φ is onto, we recall that the multiplicative system used to obtain $\big(K[X_1, \dots, X_n]/I\big)_{(P_j)}$ consists of the elements $K[X_1, \dots, X_n]/I$ that are nonzero at P_j, and φ_j carries these to units in $\big(K[X_1, \dots, X_n]/I\big)_{(P_j)}$. Since $E_j(P_j) = 1$, $\varphi_j(e_j)$ is a unit. For $i \neq j$, we have $\varphi_j(e_i)\varphi_j(e_j) = \varphi_j(e_i e_j) = 0$, and therefore $\varphi_j(e_i) = 0$. Consequently

$$\varphi_j(e_j) = \sum_{l=1}^k \varphi_j(e_l) = \varphi_j\big(\sum_{l=1}^k e_l\big) = \varphi_j(1) = 1,$$

and $\varphi_j(e_j)$ is the identity of $\big(K[X_1, \dots, X_n]/I\big)_{(P_j)}$. The localization at P_j consists of the equivalence classes of all pairs (r_j, s_j) with r_j and s_j in $K[X_1, \dots, X_n]/I$ and s_j in the multiplicative system for index j. Thus let such pairs (r_j, s_j) be given for $1 \leq j \leq k$. We are to produce an element a of $K[X_1, \dots, X_n]/I$ such that $\varphi_j(a) = \varphi_j(r_j)(\varphi_j(s_j))^{-1}$ for all j. Use of (‡) produces h_j with $h_j s_j = e_j$ for all j, and this element has the property that $\varphi_j(h_j)\varphi_j(s_j) = \varphi_j(e_j) = 1$, hence that $\varphi_j(h_j) = \varphi_j(s_j)^{-1}$. Consequently the element $a = \sum_j r_j h_j e_j$ has the property that

$$\varphi_j(a) = \varphi_j\big(\sum_i r_i h_i e_i\big) = \sum_i \varphi_j(r_i)\varphi_j(h_i)\varphi_j(e_i) = \varphi_j(r_j)(\varphi_j(s_j))^{-1}$$

and exhibits φ as onto. □

Corollary 8.13. Let K be an algebraically closed field, and let I be an ideal in the polynomial ring $K[X_1, \dots, X_n]$ whose locus of common zeros in K^n is a finite set $\{P_1, \dots, P_k\}$. Then $K[X_1, \dots, X_n]/I$ is finite-dimensional, and so is the localization $\big(K[X_1, \dots, X_n]/I\big)_{(P_j)}$ for each j.

PROOF. This is a corollary partly of the statement of Theorem 8.12 and partly of the proof. Let m be as in the proof. If I_0 is the maximal ideal $(X_1, \dots, X_n)$ of $K[X_1, \dots, X_n]$, then I_0^m is the ideal generated by all monomials of degree m, and $K[X_1, \dots, X_n]/I_0^m$ is finite-dimensional. Consequently the maximal ideal $I_j = (X_1 - x_1(P_j), \dots, X_n - x_n(P_j))$ has the property that $K[X_1, \dots, X_n]/I_j^m$ is finite-dimensional. Since $I_i^m + I_j^m = K[X_1, \dots, X_n]$ for $i \neq j$, the Chinese Remainder Theorem shows that

$$K[X_1, \dots, X_n] \Big/ \bigcap_{j=1}^{k} I_j^m \cong \prod_{j=1}^{k} K[X_1, \dots, X_n]/I_j^m,$$

and the left side is therefore finite-dimensional. By $(*)$ in the proof of Theorem 8.12, $\bigcap_{j+1}^{k} I_j^m \subseteq I$, and hence $K[X_1, \dots, X_n]/I$ is finite-dimensional. Then $\big(K[X_1, \dots, X_n]/I\big)_{(P_j)}$ is finite-dimensional as a consequence of the statement of Theorem 8.12. □

PROOF OF THEOREM 8.10e. If F and G have a common factor H of degree ≥ 1 such that $H(P) = 0$, we may assume that H is irreducible. Introduce affine local coordinates about P. If f, g, h denote the local versions of F, G, H, then the ideal (f, g) of $K[X, Y]$ is contained in the principal ideal (h). The latter ideal is proper because $h(0, 0) = 0$, and the irreducibility of H thus implies that (h) is prime. If S denotes the multiplicative system in $K[X, Y]$ of polynomials that are nonvanishing at $(0, 0)$, then $S^{-1}(f, g) \subseteq S^{-1}(h)$, and we have a natural quotient homomorphism of $S^{-1}K[X, Y]/S^{-1}(f, g)$ onto $S^{-1}K[X, Y]/S^{-1}(h)$. The latter is isomorphic as a K algebra to $(K[X, Y]/(h))_{(0,0)}$, and the dimension of this localization is a lower bound for $I(P, F \cap G)$. Since $K[X, Y]/(h)$ is an integral domain, $K[X, Y]/(h)$ maps one-one into any localization of itself, and $\dim_K(K[X, Y]/(h))$ is a lower bound for $I(P, F \cap G)$. Since h is nonconstant, either X or Y actually occurs in it, say Y. Then h divides no member of $K[X]$, and the mapping of $K[X]$ into cosets modulo (h) is one-one. Therefore $K[X, Y]/(h)$ contains a subalgebra isomorphic to $K[X]$ and must be infinite-dimensional.

Conversely if F and G have no common factor of degree ≥ 1 with P on its locus, then (d) shows that we may assume F and G to have no common factor of degree ≥ 1 of any kind. In this case Theorem 8.5 shows that the locus of common zeros of F and G is finite, and Corollary 8.13 shows that $I(P, F \cap G)$ is finite. □

PROOF OF THEOREM 8.10f. We are to prove that

$$I(P, F \cap GH) = I(P, F \cap G) + I(P, F \cap H). \qquad (*)$$

If F and GH have a common factor of degree ≥ 1 that vanishes at P, then F and one of G and H have such a factor. By symmetry we may assume that F and G

have that common factor. Then the left side of $(*)$ and the first term on the right are infinite by (e), and $(*)$ is verified.

Thus we may assume that F and GH have no common factor that vanishes at P. If F has a prime factor that does not vanish at P, then (d) shows that we can drop that factor from all three appearances of F in $(*)$. In other words, it is enough to prove (f) under the assumption that F and GH have no common factor of degree ≥ 1 of any kind.

With this assumption in place, introduce affine local coordinates about P, let S denote the multiplicative system in $K[X, Y]$ of polynomials that are nonvanishing at $(0, 0)$, and let f, g, h be the local versions of the given curves F, G, H. The inclusion of ideals $(f, gh) \subseteq (f, g)$ induces an inclusion $S^{-1}(f, gh) \subseteq S^{-1}(f, g)$ and then an onto algebra homomorphism

$$\varphi : S^{-1}K[X, Y]/S^{-1}(f, gh) \to S^{-1}K[X, Y]/S^{-1}(f, g).$$

We shall exhibit a K vector-space isomorphism ψ of $S^{-1}K[X, Y]/S^{-1}(f, h)$ onto $\ker\varphi$, and the resulting dimensional equality

$$\begin{aligned}&\dim_K \big(S^{-1}K[X, Y]/S^{-1}(f, gh)\big)\\ &\quad= \dim_K \big(S^{-1}K[X, Y]/S^{-1}(f, g)\big) + \dim_K \big(S^{-1}K[X, Y]/S^{-1}(f, h)\big) \qquad (**)\end{aligned}$$

will prove $(*)$ and hence (f). We define

$$\Psi : S^{-1}K[X, Y] \to S^{-1}K[X, Y]/S^{-1}(f, gh)$$

as a K linear map by $\Psi(u) = gu + S^{-1}(f, gh)$. If $af + bh$ is in $S^{-1}(f, h)$, then $\Psi(af + bh) = afg + bgh + S^{-1}(f, gh) = S^{-1}(f, gh)$. Thus Ψ descends to a K linear map ψ of $S^{-1}K[X, Y]/S^{-1}(f, h)$ into $S^{-1}K[X, Y]/S^{-1}(f, gh)$. It is evident that $\varphi\Psi = 0$ and hence that $\varphi\psi = 0$, i.e., image $\psi \subseteq \ker\varphi$.

If any member $u+S^{-1}(f, gh)$ of $\ker\varphi$ is given, then $0 = \varphi(u+S^{-1}(f, gh)) = u + S^{-1}(f, g)$ shows that u is in $S^{-1}(f, g)$. Say that $u = af + bg$. Then $\psi(b+S^{-1}(f, h)) = bg+S^{-1}(f, gh) = bg+af+S^{-1}(f, gh) = u+S^{-1}(f, gh)$ shows that image $\psi \supseteq \ker\varphi$. Hence image $\psi = \ker\varphi$, i.e., ψ is onto.

To see that ψ is one-one, suppose that $\psi(u + S^{-1}(f, h))$ is the 0 coset, i.e., that $gu + S^{-1}(f, gh) = S^{-1}(f, gh)$. Then $gu = af + bgh$ with u, a, b in $S^{-1}K[X, Y]$. Clearing fractions, we may assume that u, a, b are in $K[X, Y]$. The formula $g(u - bh) = af$ in $K[X, Y]$, in the presence of the assumption that F and G have no common factor of degree ≥ 1, implies that f divides $u - bh$. Write $u - bh = cf$ with c in $K[X, Y]$. Then $u = cf + bh$, and u lies in the ideal (f, h). In other words, $u + S^{-1}(f, h)$ is the trivial coset, and ψ has been shown to be one-one. This proves $(**)$ and hence (f). □

Lemma 8.14. For any field K, let $\{L_i\}_{i\ge 1}$ be a system of nonzero homogeneous polynomials in $K[X,Y]$ of the form $L_i = a_iX + b_iY$, let $\{M_j\}_{j\ge 1}$ be another such system with $M_j = c_jX + d_jY$, and suppose that no L_i is a scalar multiple of some M_j. For $n \ge 1$, let $B_0, \dots, B_n$ be the system of homogeneous polynomials

$$B_k = L_1 \cdots L_k M_1 \cdots M_{n-k} \qquad \text{for } 0 \le k \le n.$$

Then $\{B_0, \dots, B_n\}$ is a vector-space basis of the space $K[X,Y]_n$ of all homogeneous polynomials in (X,Y) of degree n.

Proof. The set $\{B_0, \dots, B_n\}$ has $n+1$ elements, and $n+1$ is the dimension of $K[X,Y]_n$ because $\{X^n, X^{n-1}Y, \dots, Y^n\}$ is a basis. Thus it is enough to show that $\{B_0, \dots, B_n\}$ is linearly independent. If we have a relation $\sum_{k=0}^n c_kB_k = 0$ for scalars c_k, then we observe that L_1 divides each B_k for $k \ge 0$, and L_1 does not divide B_0 because by assumption L_1 does not divide any factor M_j. Thus $c_0 = 0$. In effect, case n of the lemma has now been reduced to case $n-1$, and the result readily follows by induction. □

Proof of Theorem 8.10g. Put $p = m_P(F)$ and $q = m_P(G)$. We pass to affine local coordinates about P, letting f and g be the members of $K[X,Y]$ corresponding to F and G. If I denotes the maximal ideal $I = (X,Y)$ in $K[X,Y]$, then f lies in I^p and g lies in I^q. We form the following sequence of K vector spaces and K linear mappings:

$$K[X,Y]/I^q \oplus K[X,Y]/I^p \xrightarrow{\psi} K[X,Y]/I^{p+q} \xrightarrow{\varphi} K[X,Y]/(I^{p+q} + (f,g)) \to 0.$$

Here the mapping φ is the algebra homomorphism induced by the inclusion $I^{p+q} \subseteq I^{p+q} + (f,g)$, and it is onto $K[X,Y]/(I^{p+q} + (f,g))$. The mapping ψ is defined by

$$\psi(a + I^q, b + I^p) = af + bg + I^{p+q}$$

and is merely K linear.

Let us see that the sequence is exact at $K[X,Y]/I^{p+q}$. Since

$$\varphi\psi(a + I^q, b + I^p) = \varphi(af + bg + I^{p+q}) = I^{p+q} + (f,g),$$

we obtain image $\psi \subseteq \ker\varphi$. If $h + I^{p+q}$ is in $\ker\varphi$, then h is in $I^{p+q} + (f,g)$, hence is of the form $u + af + bg$ with u in I^{p+q}. Then $h - u = af + bg$, and $\psi(a + I^q, b + I^p) = h - u + I^{p+q} = h + I^{p+q}$. So image $\psi \supseteq \ker\varphi$, and we have image $\psi = \ker\varphi$.

The mapping ψ descends to a one-one linear map of

$$M = (K[X,Y]/I^q \oplus K[X,Y]/I^p)\big/\ker\psi$$

into $K[X,Y]/I^{p+q}$. The vector space $K[X,Y]/I^q$ may be identified with the space of all polynomials of degree less than q, and that space is finite-dimensional. Similarly $K[X,Y]/I^p$ is finite-dimensional, and therefore

$$\dim_K M = \dim_K K[X,Y]/I^q + \dim_K K[X,Y]/I^p - \dim_K \ker\psi. \qquad (*)$$

Meanwhile, φ exhibits $K[X,Y]/(I^{p+q}+(f,g))$ as isomorphic as a vector space to $(K[X,Y]/I^{p+q})/M$. Consequently

$$\dim_K K[X,Y]/I^{p+q} = \dim_K M + \dim_K K[X,Y]/(I^{p+q}+(f,g)). \qquad (**)$$

Combining $(*)$ and $(**)$ with the simple vector-space isomorphism $K[X,Y]/I^d \cong K[X,Y,W]_{d-1}$ and with the fact from Section 3 that $\dim_K K[X,Y,W]_{d-1} = \binom{d+1}{2}$ gives

$$\begin{aligned}
\dim_K & K[X,Y]/(I^{p+q}+(f,g))\\
&= \dim_K K[X,Y]/I^{p+q} - \dim_K K[X,Y]/I^q\\
&\quad - \dim_K K[X,Y]/I^p + \dim_K \ker\psi\\
&\geq \dim_K K[X,Y]/I^{p+q} - \dim_K K[X,Y]/I^q - \dim_K K[X,Y]/I^p\\
&= \tbinom{p+q+1}{2} - \tbinom{q+1}{2} - \tbinom{p+1}{2}\\
&= pq, \qquad (\dagger)
\end{aligned}$$

with equality on the fourth line if and only if $\ker\psi = 0$.

The locus of common zeros of $I^{p+q}+(f,g)$ is just $\{0\}$, and Theorem 8.12 therefore shows that

$$\dim_K \big(K[X,Y]/(I^{p+q}+(f,g))\big)_{(0,0)} = \dim_K K[X,Y]/(I^{p+q}+(f,g)). \quad (\dagger\dagger)$$

The inclusion $(f,g) \subseteq I^{p+q}+(f,g)$ induces an algebra homomorphism of $\big(K[X,Y]/(f,g)\big)_{(0,0)}$ onto $\big(K[X,Y]/(I^{p+q}+(f,g))\big)_{(0,0)}$. Therefore

$$\dim_K \big(K[X,Y]/(f,g)\big)_{(0,0)} \geq \dim_K \big(K[X,Y]/(I^{p+q}+(f,g))\big)_{(0,0)}. \qquad (\ddagger)$$

Let S be the set-theoretic complement of $I=(X,Y)$ in $K[X,Y]$. Because of the isomorphism $S^{-1}K[X,Y]/S^{-1}J \cong \big(K[X,Y]/J\big)_{(0,0)}$ for any ideal J, equality will hold in $(\ddagger)$ if $S^{-1}(f,g) = S^{-1}(I^{p+q}+(f,g))$. Combining $(\dagger)$, $(\dagger\dagger)$, and $(\ddagger)$, we find that

$$I(P, F\cap G) \geq pq, \qquad (\ddagger\ddagger)$$

with equality if

$$I^{p+q} \subseteq S^{-1}(f, g) \qquad \text{and} \qquad \psi \text{ is one-one.} \tag{§}$$

Inequality (‡‡) completes the proof of the inequality in (g) of the theorem. Because equality holds in (‡‡) if (§) holds, we can complete the proof of all of (g) by showing that (§) holds if F and G have no tangent line in common.

Thus for the remainder of the proof, we assume that F and G have no tangent line in common. Let the tangent lines of F, repeated according to their multiplicities, be $L_1, \dots, L_p$, and let the tangent lines of G be $M_1, \dots, M_q$. Define L_i for $i > p$ to be L_p, and define M_j for $j > q$ to be M_q.

In order to prove that the first conclusion of (§), namely that $I^{p+q} \subseteq S^{-1}(f, g)$, we shall prove that $I^t \subseteq S^{-1}(f, g)$ for t sufficiently large, and then we shall prove by induction downward on t that $I^t \subseteq S^{-1}(f, g)$ as long as $t \geq p + q$. If f and g were to have a nonconstant common factor, then a tangent line for that common factor would be a tangent line for both f and g, and no such tangent line exists according to our assumption. Therefore Bezout's Theorem (Theorem 8.2) applies to f and g and shows that their locus of common zeros is finite. Let it be $\{(0, 0), Q_1, \dots, Q_l\}$. The third paragraph of the proof of Theorem 8.12 shows that there exists a polynomial h in $K[X, Y]$ such that $h(0, 0) = 1$ and $h(Q_i) = 0$ for $1 \leq i \leq l$. Then Xh and Yh vanish on $\{(0, 0), Q_1, \dots, Q_l\}$, and the Nullstellensatz (Theorem 7.1) shows that there exists N such that $(Xh)^N$ and $(Yh)^N$ lie in (f, g). Since h is in the multiplicative system S, X^N and Y^N lie in $S^{-1}(f, g)$. Any monomial of degree $\geq 2N$ contains either a factor X^N or a factor Y^N, and consequently $I^{2N} \subseteq S^{-1}(f, g)$.

Proceeding inductively downward on t, suppose that $I^t \subseteq S^{-1}(f, g)$ and that $t - 1 \geq p + q$. As in Lemma 8.14, the polynomials defined by $B_k = L_1 \cdots L_k M_1 \cdots M_{t-1-k}$ for $0 \leq k \leq t-1$ form a vector-space basis of $K[X, Y]_{t-1}$. We show that each of these lies in $S^{-1}(f, g)$; then we can conclude that $I^{t-1} \subseteq S^{-1}(f, g)$, and our induction will be complete. Let $f = f_p + f_{p+1} + \cdots$ and $g = g_q + g_{q+1} + \cdots$ be the expansions of f and g as sums of homogeneous polynomials in (X, Y). If B_k is given, then an inequality $k \geq p$ would imply that B_k contains a factor $L_1 \cdots L_p$; this is f_p up to a constant factor. An inequality $t - 1 - k \geq q$ would imply that B_k contains a factor $M_1 \cdots M_q$; this is g_q up to a constant factor. Since $k < p$ and $t-1-k < q$ would together imply the inequality $t - 1 < p + q$ that we are assuming not to be the case, one of the alternatives $k \geq p$ and $t - 1 - k \geq q$ must occur. Say the first occurs. Except for a constant factor, we then have $B_k = f_p C$ for some homogeneous polynomial $C(X, Y)$ of degree $t - 1 - p$. Substituting for f_p gives $B_k = (f - f_{p+1} - \cdots)C$. Each term $f_{p+r}C$ with $r > 0$ is of degree $(p+r) + (t - 1 - p) > t - 1$ and therefore lies in $I^t \subseteq S^{-1}(f, g)$. Also, the term fC lies in $S^{-1}(f, g)$. Hence B_k lies in $S^{-1}(f, g)$. This completes the induction, and we conclude that $I^{p+q} \subseteq S^{-1}(f, g)$.

In order to prove the second conclusion of (§), namely that ψ is one-one, suppose that $0 = \psi(a + I^q, b + I^p) = af + bg + I^{p+q}$, i.e., that all terms of $af + bg$ are of order $\geq p + q$. Write $a = a_r + a_{r+1} + \cdots$ with $a_r \neq 0$ if a is not in I^q, and write $b = b_s + b_{s+1} + \cdots$ with $b_s \neq 0$ if b is not in I^p, so that

$$af + bg = a_r f_p + b_s g_q + \text{(higher-order terms)}.$$

The right side is assumed to be in I^{p+q}, which means that one of the following two conditions is satisfied:

(i) $r + p = s + q < p + q$ and $a_r f_p + b_s g_q = 0$,
(ii) $a_r f_p$ is in I^{p+q}, and $b_s g_q$ is in I^{p+q}.

If (i) holds, then the facts that $a_r f_p = -b_s g_q$ and that f and g have no tangent lines in common imply that f_p divides b_s. Since $s < p$, we must have $b_s = 0$. Therefore $a_r = 0$, and the conditions on a_r and b_s imply that a is in I^q and b is in I^p, which we are trying to show. If (ii) holds, then the fact that $a_r f_p$ is in I^{p+q} implies that $a_r = 0$ or $r \geq q$; in either case, a is in I^q. Similarly the fact that $b_s g_q = 0$ implies that $b_s = 0$ or $s \geq p$; in either case, b is in I^p. We conclude that ψ is one-one, as was to be shown. □

6. General Form of Bezout's Theorem for Plane Curves

With the discussion complete concerning intersection multiplicity for general projective plane curves, we arrive at the general form of Bezout's Theorem for plane curves.

Theorem 8.15 (Bezout's Theorem). Let K be an algebraically closed field, and let F and G be projective plane curves over K of respective degrees m and n. If F and G have no common factor of positive degree, then

$$\sum_{P \in \mathbb{P}^2_K} I(P, F \cap G) = mn.$$

REMARKS. The sum over P has only finitely many nonzero terms by Theorem 8.5, and each intersection multiplicity in the sum is finite by Theorem 8.10e.

PROOF. Theorem 8.5 shows that the locus of common zeros of F and G is a finite set. By applying a suitable Φ in GL$(3, K)$, we may assume that none of these zeros lies on the line at infinity, namely W. To do so, we choose a point P not in the finite set of common zeros. There are only finitely many lines passing through P and some member of the set of common zeros, and we choose a line through P different from all these. If Φ is chosen so as to move this line to the line at infinity W, then none of the common zeros will lie on the line W.

With this normalization in place, let $\{P_1, \dots, P_k\}$ be the set of common zeros of F and G. We introduce local versions f and g of F and G by the definitions $f(X, Y) = F(X, Y, 1)$ and $g(X, Y) = G(X, Y, 1)$. Application of Theorem 8.12 to the ideal $I = (f, g)$ in $K[X, Y]$ gives

$$\dim_K K[X, Y]/(f, g) = \sum_{j=1}^{k} \dim_K \big(K[X, Y]/(f, g)\big)_{(P_j)} = \sum_{j=1}^{k} I(P_j, F \cap G).$$

The theorem will therefore follow if we prove that

$$\dim_K K[X, Y]/(f, g) = mn. \tag{$*$}$$

To prove $(*)$, we shall first prove a related equality concerning $K[X, Y, W]$ and the ideal (F, G) in it, and then we shall use the fact that F and G have no common zeros with W to transfer the conclusion to $K[X, Y]$.

Define K linear mappings $\varphi : K[X, Y, W] \oplus K[X, Y, W] \to K[X, Y, W]$ and $\psi : K[X, Y, W] \to K[X, Y, W] \oplus K[X, Y, W]$ by

$$\varphi(A, B) = AF + BG \qquad \text{and} \qquad \psi(C) = (CG, -CF),$$

and form the sequence of K vector spaces and K linear maps given by

$$0 \longrightarrow K[X, Y, W] \xrightarrow{\psi} K[X, Y, W] \oplus K[X, Y, W] \xrightarrow{\varphi} K[X, Y, W]. \tag{$**$}$$

It is evident that ψ is one-one, that $\varphi\psi = 0$, and that image $\varphi = (F, G)$. If (A, B) is in $\ker \varphi$, then $AF + BG = 0$. Since F and G have no common factor of positive degree, F divides B and G divides A. Setting $C = AG^{-1}$ therefore gives $A = CG$ and $B = -AG^{-1}F = -CF$. Hence (A, B) lies in image ψ. In other words, $(**)$ is exact, and image $\varphi = (F, G)$.

Let $d \geq m + n$. If we denote by ψ_d and φ_d the restrictions of ψ and φ to $K[X, Y, W]_{d-m-n}$ and $K[X, Y, W]_{d-n} \oplus K[X, Y, W]_{d-m}$, respectively, and if we go over the argument in the previous paragraph, then we see that the sequence

$$0 \longrightarrow K[X,Y,W]_{d-m-n} \xrightarrow{\psi_d} K[X,Y,W]_{d-n} \oplus K[X,Y,W]_{d-m} \xrightarrow{\varphi_d} K[X,Y,W]_d$$

is exact and that image $\varphi_d = (F, G)_d$. The vector spaces in question here are all finite-dimensional, and thus we obtain

$$\begin{aligned}
&\dim_K (F, G)_d \\
&\quad = \dim_K K[X, Y, W]_{d-n} + \dim_K K[X, Y, W]_{d-m} - \dim_K K[X, Y, W]_{d-m-n} \\
&\quad = \tbinom{d-n+2}{2} + \tbinom{d-m+2}{2} - \tbinom{d-m-n+2}{2} \\
&\quad = -mn + \tbinom{d+2}{2} \\
&\quad = -mn + \dim_K K[X, Y, W]_d. \qquad\qquad (\dagger)
\end{aligned}$$

The ideal (F, G) is homogeneous, and thus we know from Section 3 that the image of $K[X, Y, W]_d$ in $K[X, Y, W]/(F, G)$ is $K[X, Y, W]_d/(F, G)_d$. If we write $\big(K[X, Y, W]/(F, G)\big)_d$ for this quotient, then (†) shows that

$$\dim_K \big(K[X, Y, W]/(F, G)\big)_d = mn \qquad (\dagger\dagger)$$

for all $d \geq m + n$.

To prove $(*)$ and the theorem, we shall translate $(\dagger\dagger)$ into a conclusion about $K[X, Y]/(f, g)$. Fix $d \geq m + n$, and let $\{V_1 + (F, G), \dots, V_{mn} + (F, G)\}$ be a K basis of $\big(K[X, Y, W]/(F, G)\big)_d$. Define $v_j(X, Y) = V_j(X, Y, 1)$ for each j. We shall prove that the vectors

$$v_1 + (f, g), \dots, v_{mn} + (f, g) \qquad (\ddagger)$$

form a K basis of $K[X, Y]/(f, g)$.

We need to make use of the fact that F and G have no common zeros on the line at infinity. Since $W(F, G) \subseteq (F, G)$, the K linear mapping of multiplication by W on $K[X, Y, W]$ descends to a K linear mapping L of $K[X, Y, W]/(F, G)$ to itself defined by $L(H + (F, G)) = WH + (F, G)$. Let us see that

$$L : K[X, Y, W]/(F, G) \to K[X, Y, W]/(F, G) \quad \text{is one-one.} \qquad (\ddagger\ddagger)$$

In fact, suppose that $WH = AF + BG$ for some H in $K[X, Y, W]$. For any U in $K[X, Y, W]$, let $U_0(X, Y) = U(X, Y, 0)$. If U is homogeneous, then so is U_0. In this notation we can write $F = F_0 + WM$ and $G = G_0 + WN$ for homogeneous members M and N of $K[X, Y, W]$. The polynomials F_0 and G_0 are relatively prime: in fact, if F_0 and G_0 have a nontrivial common factor D_0, then we can regard D_0 as a projective plane curve, and it must have a common zero Q with W, by Theorem 8.5; but then F, G, and W have Q as a common zero, in contradiction to the normalization in the first paragraph of the proof. Since $WH = AF + BG$ implies $A_0F_0 = -B_0G_0$, it follows that F_0 divides B_0 and that G_0 divides A_0. In other words, $B_0 = C_0F_0$ and $A_0 = -C_0G_0$ for some C_0 in $K[X, Y]$. If we define $A' = A + C_0G$ and $B' = B - C_0F$, then the formulas for A_0 and B_0 show that $A'_0 = B'_0 = 0$. Hence $A' = WA''$ and $B' = WB''$ for some homogeneous polynomials A'' and B''. Then $WH = AF + BG = (A' - C_0G)F + (B' + C_0F)G = A'F + B'G = W(A''F + B''G)$, and we obtain $H = A''F + B''G$. Thus H lies in (F, G), and $(\ddagger\ddagger)$ is proved.

Left multiplication L by W carries $K[X, Y, W]_d$ into $K[X, Y, W]_{d+1}$ and carries $(F, G)_d$ into $(F, G)_{d+1}$. Therefore L is well defined as a mapping from $\big(K[X, Y, W]/(F, G)\big)_d$ into $\big(K[X, Y, W]/(F, G)\big)_{d+1}$. Since it is one-one by $(\ddagger\ddagger)$ and since the spaces are finite-dimensional, it is onto. Therefore

$$\{W^r V_1 + (F, G), \dots, W^r V_{mn} + (F, G)\} \qquad \text{is a basis} \qquad (\S)$$

of $\big(K[X, Y, W]/(F, G)\big)_{d+r}$ for every $r \geq 0$.

To prove that (‡) spans $K[X, Y]/(f, g)$, let h be in $K[X, Y]$. Let H be a homogeneous polynomial in $K[X, Y, W]$ with $h(X, Y) = H(X, Y, 1)$, and choose an integer s such that $W^s H$ lies in $K[X, Y, W]_{d+r}$ for some $r \geq 0$. Then we can write $W^s H = \sum_{j=1}^{mn} c_j W^r V_j + AF + BG$ for suitable scalars c_j and homogeneous polynomials A and B. Restricting the domain to points $(X, Y, 1)$ gives $h = \sum_{j=1}^{mn} c_j v_j + af + bg$, and therefore $h + (f, g) = \sum_{j=1}^{mn} c_j v_j + (f, g)$. This proves that (‡) spans $K[X, Y]/(f, g)$.

To prove that (‡) is linearly independent, suppose that $\sum_{j=1}^{mn} c_j v_j = af + bg$ with a and b in $K[X, Y]$. If A and B are homogeneous polynomials such that $a(X, Y) = A(X, Y, 1)$ and $b(X, Y) = B(X, Y, 1)$, then $W^r \sum_{j=1}^{mn} c_j V_j = W^s AF + W^t BG$, provided the exponents r, s, t are chosen to make the degrees of the terms $W^r \sum_{j=1}^{mn} c_j V_j$, $W^s AF$, and $W^t BG$ match. Consequently $W^r \sum_{j=1}^{mn} c_j V_j$ lies in $(F, G)_{d+r}$, and (§) shows that the coefficients are all 0. This proves that (‡) is linearly independent. □

7. Gröbner Bases

The remainder of the chapter returns to the main question introduced in Section 1, that of how to get information about the set of simultaneous solutions of polynomial equations in several variables. The resultant introduced in Section 2 gave us one tool, but the tool is of most use when there are only two equations. Beyond two equations the number of cases to check quickly grows, and the resultant is of limited usefulness.[12]

The tool to be introduced in this section is of a completely different nature. Historically it was introduced in order to have a way of deciding whether an ideal in $K[X_1, \ldots, X_n]$ contains a given polynomial. We know from the Hilbert Basis Theorem that every such ideal is finitely generated, and it is assumed that the ideal to be tested is specified by such a set of generators.

The proof of the Hilbert Basis Theorem gives a clue how to start studying an ideal of polynomials. In the statement of the theorem, R is a Noetherian integral domain, and I is a nonzero ideal in $R[X]$. It is to be proved that I is finitely generated. The proof by Hilbert is longer than the proof given in *Basic Algebra*, but the idea is clearer. To each nonzero member $f(X)$ of I, we associate the coefficient of the highest power of X appearing in $f(X)$. These coefficients, together with 0, form an ideal $L(I)$ in R, and $L(I)$ is finitely generated because R is Noetherian. Let $a_1, \ldots, a_r$ be generators, let $f_1(X), \ldots, f_r(X)$ be members

[12] The nature of the extended theory can be found in Van der Waerden, Volume II, Chapter XI. Theorem 8.31 below in effect reproduces some of this extended theory in a context that is manageable because of the theory of Gröbner bases.

of I with respective highest coefficients $a_1, \dots, a_r$, and let q be the largest of the degrees of $f_1(X), \dots, f_r(X)$. If a general $g(X)$ in I is given and if $a \in R$ is its highest coefficient, then we know that $a = \sum_i c_i a_i$ with $c_i \in R$. The polynomial $h(X)$ given by $h(X) = g(X) - \sum_i c_i f_i(X) X^{\deg g - \deg f_i}$ has degree lower than $\deg g$, and $g(X)$ will be in $(f_1, \dots, f_r)$ if $h(X)$ is in $(f_1, \dots, f_r)$. Iterating this construction, we see that it is enough to account for all the members of I of degree $\leq q - 1$. To handle these, one way to proceed is to enlarge the set $\{f_1, \dots, f_r\}$ a little. For each k with $0 \leq k \leq q - 1$, let $L_k(I)$ be the union of $\{0\}$ and the set of coefficients of X^k in members of I of degree k. Each of these is an ideal of R and hence is finitely generated, and we adjoin to $\{f_1, \dots, f_r\}$ a finite set of generators for each $L_k(I)$ with $0 \leq k \leq q - 1$. The result is a finite set $\{g_1, \dots, g_s\}$ of generators of I, as one easily checks.

In fact, the set $\{g_1, \dots, g_s\}$ is a special set of generators. For any member f of $R[X]$, let $\mathrm{LT}(f)$ be the complete term of $f(X)$ containing the highest power of X. What the argument shows is that $\{g_1, \dots, g_s\}$ is a subset of I such that $\mathrm{LT}(I) = \big(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_s)\big)$, where $\mathrm{LT}(I)$ denotes the ideal given as the linear span of all polynomials $\mathrm{LT}(g)$ for g in I. One can show that this property of $\{g_1, \dots, g_s\}$ implies that $\{g_1, \dots, g_s\}$ generates I. In essence this property will be the defining property of a "Gröbner basis" of I. It is not automatically satisfied for just any finite generating set $\{f_1, \dots, f_r\}$, as the example below shows. We shall see that it is easy to use such a set of generators to test any polynomial in $R[X]$ for membership in I. Thus the original problem historically for introducing such sets is solved except for one little detail: the proof of the Hilbert Basis Theorem is not constructive, and we are left with no idea how actually to construct a Gröbner basis.[13]

EXAMPLE. Treat $K[X, Y]$ as an instance of the above setting by letting $R = K[Y]$ and regarding $K[X, Y]$ as $R[X]$. Consider the ideal $I = (f_1, f_2)$ in $R[X]$ with $f_1(X, Y) = X^2 + 2XY^2$ and $f_2(X, Y) = XY + 2Y^3 - 1$. Then $\big(\mathrm{LT}(f_1), \mathrm{LT}(f_2)\big) = (X^2, XY)$, and every monomial appearing with nonzero coefficient in a member of the latter ideal has total degree at least 2. On the other hand, I contains the polynomial

$$Yf_1(X, Y) - Xf_2(X, Y) = Y(X^2 + 2XY) - X(XY + 2Y^3 - 1) = X,$$

and its leading term is X, whose total degree is 1. Thus $\mathrm{LT}(I)$ properly contains $\big(\mathrm{LT}(f_1), \mathrm{LT}(f_2)\big)$.

Because of the nonconstructive nature of the proof of the Hilbert Basis Theorem, it is necessary to start afresh. One message to glean from the abstract proof

[13] The exposition in this section and the next three is based partly on the book of Cox–Little–O'Shea and a now-defunct Web tutorial of Fabrizio Caruso.

is that the leading terms of the members of I are important and somewhat control the nature of I. To handle $K[X_1, \dots, X_n]$ when K is a field, it is of course necessary to use an additional induction that enumerates the variables. In the example above, we treated X as more significant than Y. For the inductive step for general $K[X_1, \dots, X_n]$, the ring R in the above argument is K with some number m of the indeterminates included, and X is the $(m+1)^{\text{st}}$ indeterminate. Putting all the steps of the induction together, we see that the order in which the variables are processed appears to be important.

The theory of Gröbner bases as it has evolved allows a healthy extra measure of generality. Instead of defining leading terms by insisting on an ordering of the indeterminates, it defines them by using a suitable kind of ordering of monomials, and that is where we begin. Let $K[X_1, \dots, X_n]$ be given, K being a field. Let $\mathcal{M}$ be the set of all monomials in $K[X_1, \dots, X_n]$. A **monomial ordering** $\leq$ on $\mathcal{M}$ is a total ordering[14] with the two additional properties that

(i) $M_1 \leq M_2$ implies $M_1M_3 \leq M_2M_3$ for all M_1, M_2, M_3 in $\mathcal{M}$,

(ii) $1 \leq M$ for all M in $\mathcal{M}$.

We write $M_2 \geq M_1$ to mean $M_1 \leq M_2$. Also, $M_1 < M_2$ means $M_1 \leq M_2$ with $M_1 \neq M_2$, and $M_1 > M_2$ means $M_1 \geq M_2$ with $M_1 \neq M_2$.

EXAMPLES OF MONOMIAL ORDERINGS. Each ordering assumes that the variables are enumerated in some way. In these examples we take this enumeration to be $X_1, \dots, X_n$. The first four examples all have the property that the largest X_j is X_1 and the smallest is X_n.

(1) **Lexicographic ordering**, abbreviated as "lex" by many authors and written as $\leq_{\text{LEX}}$ in this list of examples. This, the most important monomial ordering, is already suggested by the proof of the Hilbert Basis Theorem. In principle it can be used for all purposes in Sections 7–10, but one application in Chapter X will require a different monomial ordering. Its disadvantage is that it sometimes makes lengthy computations take longer than necessary; this matter will be discussed more in Section 9. The definition is that $X_1^{i_1} \cdots X_n^{i_n} \leq_{\text{LEX}} X_1^{j_1} \cdots X_n^{j_n}$ if either the two monomials are equal or else the first k for which $i_k \neq j_k$ has $i_k < j_k$. Thus for example, $X_1X_2^2X_3^3 \leq_{\text{LEX}} X_1^2$. The word "lexicographic" refers to the dictionary system for alphabetizing in which a first word comes before a second word if for the first position in which the two words differ, the letter of the first word in that position precedes alphabetically the letter of the second word in that position.

(2) **Graded lexicographic ordering**, abbreviated as "glex" or "grlex" by many authors. As in Section 3 the **total degree** of a monomial $X_1^{i_1} \cdots X_n^{i_n}$ is

[14] This means a partial ordering with the properties that each pair a, b has $a \leq b$ or $b \leq a$ and that both hold only if $a = b$.

$\deg(X_1^{i_1}\cdots X_n^{i_n}) = \sum_{k=1}^n i_k$. The definition of the ordering is that $M \le_{\text{GLEX}} N$ if either $\deg M < \deg N$ or else if $\deg M = \deg N$ and $M \le_{\text{LEX}} N$. Thus for example, $X_1^2 \le_{\text{GLEX}} X_1X_2^2X_3^3$ because the total degree 2 of the first monomial is less than the total degree 6 of the second monomial. But $X_1X_2^2X_3^3 \le_{\text{GLEX}} X_1^2X_3^4$ because both monomials have the same total degree 6 and the second monomial involves a higher power of X_1 than does the first. This monomial ordering is not much used; more common is the variant of it in the next example.

(3) **Graded reverse lexicographic ordering**, abbreviated as "grevlex" by many authors. The definition is that $M \le_{\text{GREVLEX}} N$ if either $\deg M < \deg N$ or else if $\deg M = \deg N$ and $N^t \le_{\text{LEX}} M^t$, where M^t is M but with the exponents of X_j and X_{n-j} interchanged for each j, and where N^t is defined similarly. This ordering takes some getting used to. For example, $X_1^2X_3^4 \le_{\text{GREVLEX}} X_1X_2^2X_3^3$ when $n = 3$ because both monomials have the same total degree and $X_1^3X_2^2X_3 = (X_1X_2^2X_3^3)^t \le_{\text{LEX}} (X_1^2X_3^4)^t = X_1^4X_3^2$. By contrast, $X_1X_2^2X_3^3 \le_{\text{GLEX}} X_1^2X_3^4$.

(4) Orderings of k-**elimination type**, where $1 \le k \le n-1$. These are orderings such that any monomial containing one of $X_1, \dots, X_k$ to a positive power exceeds any monomial in $X_{k+1}, \dots, X_n$ alone. These will be discussed in Section 10. Of them, one of particular importance is the **Bayer–Stillman ordering** of k-elimination type. Here a monomial M is $\le$ a monomial N if the sum of the exponents of $X_1, \dots, X_k$ for M is less than the corresponding sum for N or else the two sums are equal and $M \le_{\text{GREVLEX}} N$. This ordering is commonly used for making computations in the context of Section 10.

(5) Ordering from a tuple of **weight vectors**. For $1 \le i \le n$, let $w^{(i)}$ be a vector in $\mathbb{R}^n$ of the form $w^{(i)} = (w_1^{(i)}, \dots, w_n^{(i)})$, and assume that $w^{(1)}, \dots, w^{(n)}$ are linearly independent over $\mathbb{R}$. Identify the monomial X^α with the vector of individual exponents $\alpha = (\alpha_1, \dots, \alpha_n)$. The ordering given by the weight vectors $w_j^{(i)}$ is defined by saying that $X^\alpha \le X^\beta$ if $X^\alpha = X^\beta$ or if the first i such that $w^{(i)} \cdot \alpha \ne w^{(i)} \cdot \beta$ has $w^{(i)} \cdot \alpha < w^{(i)} \cdot \beta$. Here the dot refers to the ordinary dot product. A condition is needed on the $w^{(i)}$'s to ensure that $1 \le X^\alpha$ for all α. (See Problem 14 at the end of the chapter.) Here are two specific examples for which the condition is satisfied. Let $e^{(i)}$ be the i^{th} standard basis vector of $\mathbb{R}^n$. The lexicographic ordering in Example 1 is determined by the tuple of weight vectors $(e^{(1)}, \dots, e^{(n)})$. The Bayer–Stillman ordering in Example 4 is determined by the tuple of weight vectors

$$\big(e^{(1)} + \cdots + e^{(k)}, e^{(k+1)} + \cdots + e^{(n)}, -e^{(n)}, \dots, -e^{(k+2)}, -e^{(k)}, \dots, -e^{(2)}\big).$$

Further discussion of monomial orderings determined by weight vectors occurs in Problems 14–15 at the end of the chapter.

Property (i) of monomial orderings insists that the ordering respect multiplication of monomials in the natural way. Property (ii), according to the next proposition, is a well-ordering property. The proof of the proposition will be preceded by a lemma.

Proposition 8.16. In any monomial ordering for $K[X_1, \dots, X_n]$, any decreasing sequence $M_1 \geq M_2 \geq M_3 \geq \cdots$ is eventually constant. Consequently each nonempty subset of $\mathcal{M}$ has a smallest element in the ordering.

Lemma 8.17. If I is an ideal in $K[X_1, \dots, X_n]$ generated by monomials and if $f(X_1, \dots, X_n)$ is in I, then each monomial appearing in the expansion of f with nonzero coefficient lies in I. Consequently I has a finite set of monomials as generators. Moreover, if $\{M_1, \dots, M_s\}$ is a set of monomials that generate I and if M is any monomial in I, then some M_j divides M.

PROOF. Let $\{M_\alpha\}$ be the set of monomials that generates I. If f is in I, then we can write $f = \sum_{j=1}^k h_j M_{\alpha_j}$ for polynomials h_j. Let $h_j = \sum_{i=1}^{l_j} c_{ij} M_{ij}$ be the expansion of h_j in terms of monomials. If M_0 is a monomial appearing in f with nonzero coefficient c, then the only possible monomial M_{ij} in h_j that can contribute toward c is one with $M_{ij} M_{\alpha_j} = M_0$ if such a monomial exists. For some j, such a monomial must exist, or c would be 0; thus M_0 lies in I.

For the second conclusion, write $\{f_1, \dots, f_l\}$ by the Hilbert Basis Theorem. The first conclusion shows that each monomial contributing to each f_j lies in I, and the set of all these monomials, as j varies, is therefore a finite set of monomials generating I.

For the third conclusion, write $M = \sum_{i=1}^s a_i M_i$ for polynomials a_i. Expanding each a_i in terms of monomials, we see that some a_i contains with nonzero coefficient a monomial M' such that $M = M' M_i$. The divisibility follows. □

PROOF OF PROPOSITION 8.16. Let M be a monomial, and let I be the linear span of all monomials M' with $M' \geq M$. If M' is a such a monomial and N is any monomial, then $NM' \geq NM$ by (i), and $NM \geq 1M = M$ by (i) and (ii). Therefore NM' lies in I, and I is an ideal.

From such an ideal I, we can recover M as the unique monomial M_0 in I such that $M_0 \leq M'$ for every monomial M' in I, since any such M_0 has $M_0 \leq M$ as well as $M \leq M_0$.

With $M_1, M_2, \dots$ given as in the proposition, let I_k be the linear span of all monomials $M' \geq M_k$. We have just seen that I_k is an ideal, and the I_k's are increasing in k. Then $I = \bigcup_{k=1}^\infty I_k$ is an ideal generated by monomials, and Lemma 8.17 shows that it has a finite set of monomials as a set of generators. Each such monomial generator lies in some I_k. Since the I_k's are nested, all the generators lie in some I_{k_0}, and we conclude that $I = I_{k_0}$. The previous paragraph of the proof shows that I_{k_0} determines M_{k_0}, and therefore $M_k = M_{k_0}$ for all $k \geq k_0$.

For the last statement of the proposition, if there were no least element, then for any element in the subset, we could always find a smaller element in the subset. In this way, we would be able to construct a strictly decreasing infinite sequence in $\mathcal{M}$, in contradiction to what has just been proved. □

Fix a monomial ordering for $K[X_1, \dots, X_n]$. If f is any nonzero member of $K[X_1, \dots, X_n]$ and if f is expanded as a K linear combination of monomials, then we define the leading monomial, leading coefficient, and leading term of f by

$$\begin{aligned} \text{LM}(f) &= \text{largest monomial with nonzero coefficient in expansion of } f, \\ \text{LC}(f) &= \text{coefficient of } \text{LM}(f) \text{ in } f, \\ \text{LT}(f) &= \text{LC}(f)\,\text{LM}(f). \end{aligned}$$

It will be convenient to be able to use these definitions without having to distinguish the cases $f \neq 0$ and $f = 0$. Accordingly, let us adjoin 0 to the set $\mathcal{M}$, agreeing that $0 < M$ and $0M = 0$ for every monomial M. We adopt the convention that $\text{LM}(0) = 0$, $\text{LT}(0) = 0$, and $\text{LC}(0) = 0$.

Since any monomial that occurs in a sum of two polynomials occurs in one or the other of them, it is immediate from the definition that

$$\text{LM}(f_1 + f_2) \le \max(\text{LM}(f_1), \text{LM}(f_2))$$

if f_1, f_2, and $f_1 + f_2$ are nonzero. Checking the various cases, we see that this inequality persists if one or more of f_1, f_2, and $f_1 + f_2$ are 0.

The comparable results concerning multiplication are contained in the next proposition.

Proposition 8.18. If f_1 and f_2 are two nonzero members of $K[X_1, \dots, X_n]$, then

$$\text{LM}(f_1 f_2) = \text{LM}(f_1)\,\text{LM}(f_2) \qquad \text{and} \qquad \text{LC}(f_1 f_2) = \text{LC}(f_1)\,\text{LC}(f_2);$$

hence

$$\text{LT}(f_1 f_2) = \text{LT}(f_1)\,\text{LT}(f_2).$$

These equalities persist if one or both of f_1 and f_2 are 0. Moreover, if f_1 and f_2 are nonzero and have $\text{LT}(f_1) = \text{LT}(f_2)$, then $\text{LM}(f_1 - f_2) < \text{LM}(f_1)$.

PROOF. For the first statement, let the expansions of f_1 and f_2 as linear combinations of distinct monomials be $f_1 = a_1\,\text{LM}(f_1) + \sum_i c_i M_i$ and $f_2 = a_2\,\text{LM}(f_2) + \sum_j d_j N_j$ with $M_i < \text{LM}(f_1)$ for all i and $N_j < \text{LM}(f_2)$ for all j. Then $f_1 f_2$ equals

$$a_1 a_2\,\text{LM}(f_1)\,\text{LM}(f_2) + a_2 \sum_i c_i M_i\,\text{LM}(f_2) + a_1 \sum_j d_j\,\text{LM}(f_1) N_j + \sum_{i,j} c_i d_j M_i N_j,$$

and the conclusions in the first sentence of the proposition will follow if it is shown that $M_i\,\mathrm{LM}(f_2) < \mathrm{LM}(f_1)\,\mathrm{LM}(f_2)$, that $\mathrm{LM}(f_1)N_j < \mathrm{LM}(f_1)\,\mathrm{LM}(f_2)$, and that $M_iN_j < \mathrm{LM}(f_1)\,\mathrm{LM}(f_2)$. The first inequality follows from (i) because $M_i < \mathrm{LM}(f_1)$, and the second inequality is similar. For the third we apply (i) twice to obtain $M_iN_j \le M_i\,\mathrm{LM}(f_2) \le \mathrm{LM}(f_1)\,\mathrm{LM}(f_2)$ and observe that the end expressions can be equal only if equality holds in both instances. The latter is impossible because $K[X_1, \dots, X_n]$ is an integral domain, and thus $M_iN_j < \mathrm{LM}(f_1)\,\mathrm{LM}(f_2)$.

The three displayed equalities persist if one or both of f_1 and f_2 are 0 because $\mathrm{LM}(f)$, $\mathrm{LT}(f)$, and $\mathrm{LC}(f)$ can be 0 only if $f = 0$.

Finally if f_1 and f_2 are nonzero and have expansions as in the first paragraph of the proof with $\mathrm{LT}(f_1) = \mathrm{LT}(f_2)$, then $\mathrm{LC}(f_1) = a_1$ and $\mathrm{LC}(f_2) = a_2$. Hence $f_1 - f_2$ has an expansion involving only the monomials M_i and N_j. Consequently if $f_1 - f_2 \ne 0$, then the largest of the M_i's and N_j's is $< \mathrm{LM}(f_1)$. Thus $\mathrm{LM}(f_1 - f_2) < \mathrm{LM}(f_1)$. This inequality holds also if $f_1 - f_2 = 0$. □

If I is a nonzero ideal in $K[X_1, \dots, X_n]$, we define $\mathrm{LT}(I)$ to be the vector space of all K linear combinations of polynomials $\mathrm{LT}(f)$ with f in I. It follows from Proposition 8.18 that $K[X_1, \dots, X_n]\,\mathrm{LT}(I) \subseteq \mathrm{LT}(I)$, and therefore $\mathrm{LT}(I)$ is an ideal in $K[X_1, \dots, X_n]$. A finite unordered subset $\{g_1, \dots, g_k\}$ of nonzero elements of the ideal I is called a **Gröbner basis** of I if $\mathrm{LT}(I) = \big(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_k)\big)$. The inclusion $\supseteq$ follows from the definition, and the question is whether $\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_k)$ generate $\mathrm{LT}(I)$.

Among the examples below, Example 3 is particularly suggestive of the utility of a Gröbner basis. The idea is that an ordinary set of generators may have the property that certain "small" elements of I can be expanded in terms of the generators only using "large" coefficients and that this property is reflected in the failure of $(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_k))$ to exhaust $\mathrm{LT}(I)$.

EXAMPLES WITH LEXICOGRAPHIC ORDERING.

(1) Principal ideal. If $I = (f(X_1, \dots, X_n))$, then $\{f\}$ is a Gröbner basis. In fact, the most general member of I is of the form hf with h in $K[X_1, \dots, X_n]$, and Proposition 8.18 gives $\mathrm{LT}(hf) = \mathrm{LT}(h)\,\mathrm{LT}(f)$. Therefore $\mathrm{LT}(I) = (\mathrm{LT}(f))$, as required.

(2) Ideal generated by members of $K[X_1, \dots, X_n]_1$. Suppose that $I = (L_1, \dots, L_k)$, where each L_j is a homogeneous linear polynomial of degree 1. For example, I could be $(X_1 + X_2 + X_3, X_1 - X_3)$. Let us form the corresponding k-by-n coefficient matrix, specifically $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{pmatrix}$ in the 3-variable example. If we perform row operations to transform this matrix into reduced row-echelon form and let $L'_1, \dots, L'_{k'}$ be the members of $K[X_1, \dots, X_n]_1$ corresponding to the reduced matrix, specifically $X_1 - X_3$ and $X_2 + 2X_3$ for the reduced form

$\left(\begin{smallmatrix}1&0&-1\\0&1&2\end{smallmatrix}\right)$ of $\left(\begin{smallmatrix}1&1&1\\1&0&-1\end{smallmatrix}\right)$, then $I = (L'_1, \dots, L'_{k'})$ and moreover $\{L'_1, \dots, L'_{k'}\}$ is a Gröbner basis of I. This fact is not particularly obvious in the full generality of this example, but it will be shown to be an easy consequence of Theorem 8.23 in the next section.

(3) Earlier example in this section. In $K[X, Y]$, let $I = (f_1, f_2)$ with $f_1(X, Y) = X^2 + 2XY^2$ and $f_2(X, Y) = XY + 2Y^3 - 1$. Then $\big(\mathrm{LT}(f_1), \mathrm{LT}(f_2)\big) = (X^2, XY)$. We saw that X is a member of I and that $\mathrm{LT}(X) = X$ is not in $\big(\mathrm{LT}(f_1), \mathrm{LT}(f_2)\big)$. So $\{f_1, f_2\}$ is not a Gröbner basis. If we enlarge the set of generators of I to $\{f_1, f_2, X\}$, then we still do not have a Gröbner basis because $f_2 - YX = 2Y^3 - 1$ is in I and $\mathrm{LT}(f_2 - YX) = 2Y^3$ does not lie in $\big(\mathrm{LT}(f_1), \mathrm{LT}(f_2), \mathrm{LT}(X)\big) = (X^2, XY, X) = (X)$. We can enlarge the set of generators still further to $\{f_1, f_2, X, 2Y^3 - 1\}$. Is this a Gröbner basis? Here we have $\big(\mathrm{LT}(f_1), \mathrm{LT}(f_2), \mathrm{LT}(X), \mathrm{LT}(2Y^3 - 1)\big) = (X, Y^3)$, and it seems as if this equals $\mathrm{LT}(I)$. But we need a way of checking easily. We shall obtain a way of checking in Theorem 8.23 in the next section.

The question of existence–uniqueness of a Gröbner basis will be addressed constructively in Sections 8–9; however, we did observe at the beginning of this section that Hilbert's proof of the Hilbert Basis Theorem essentially handles existence when the monomial ordering is the usual lexicographic ordering. Actually, the argument at the beginning of the section had two parts to it—a nonconstructive argument producing a certain finite set of leading terms and a verification that those leading terms lead to a set of generators of the ideal. The first part, being a nonconstructive existence proof, does not help us in our current efforts, and we defer to Problem 13 at the end of the chapter the question of adapting it to a general monomial order. The second part, on the other hand, is a useful kind of verification in our current efforts. It shows that a certain kind of finite subset of an ideal is necessarily a set of generators, and it generalizes as follows. The generalization will play a role in Section 9.

Proposition 8.19. If K is a field, if a monomial ordering is specified for $K[X_1, \dots, X_n]$, and if $\{g_1, \dots, g_k\}$ is a Gröbner basis for a nonzero ideal I of $K[X_1, \dots, X_n]$, then $\{g_1, \dots, g_k\}$ generates I.

PROOF. First we prove that if $f \neq 0$ is in I, then there exist a g_j, a monomial M_0, and a nonzero scalar c such that $\mathrm{LM}(f - cM_0g_j) < \mathrm{LM}(f)$. To see this, we use the hypothesis that $\{g_1, \dots, g_k\}$ is a Gröbner basis to find polynomials $h_1, \dots, h_k$ such that $\mathrm{LM}(f) = \sum_{i=1}^k h_i \,\mathrm{LM}(g_i)$. Then it must be true for i equal to some index j that $\mathrm{LM}(f) = M_0 \,\mathrm{LM}(g_j)$ for one of the monomials M_0 that appears in h_j with nonzero coefficient. Since $M_0 \,\mathrm{LM}(g_j) = \mathrm{LM}(M_0)\,\mathrm{LM}(g_j) = \mathrm{LM}(M_0g_j)$, we can rewrite this equality as $\mathrm{LT}(f) = c\,\mathrm{LT}(M_0g_j)$ for some scalar $c \neq 0$. Then

$\text{LT}(f) = \text{LT}(cM_0g_j)$, and Proposition 8.18 shows that $\text{LM}(f - cM_0g_j) < \text{LM}(f)$, as asserted.

Iterating this construction and assuming that we never get 0, we can find successively nonzero scalars c_i, monomials M_i, and members g_{j_i} of the Gröbner basis such that the sequence $\text{LM}\big(f - \sum_{i=1}^{l} c_j M_j g_{j_i}\big)$ indexed by l is strictly decreasing, in contradiction to Proposition 8.16. To avoid the contradiction, we must have $f - \sum_{i=1}^{l} c_j M_j g_{j_i} = 0$ for some l, and then f is exhibited as in the ideal $(g_1, \dots, g_k)$. Hence the Gröbner basis generates I. □

8. Constructive Existence

Throughout this section, K denotes a field, and we work with a fixed monomial ordering on $K[X_1, \dots, X_n]$. Ideals in $K[X_1, \dots, X_n]$ will always be specified by giving finite sets of generators. Our objective is to obtain a constructive proof of the existence of a Gröbner basis for each nonzero ideal in $K[X_1, \dots, X_n]$, along with a useful test procedure for deciding whether a given finite set of generators of I is a Gröbner basis. As is often the case with existence proofs, the motivation for the proof comes from a certain amount of deduction of properties that a Gröbner basis must satisfy if its exists. It was mentioned in the previous section that the failure of a set of generators to be a Gröbner basis has something to do with its failure to be able to represent all "small" elements of the ideal by means of expansions in terms of the generators that use "small" coefficients. The first part of this section will explore this idea, seeking to make it precise. The main step will be a checkable text for a set to be a Gröbner basis; this is Theorem 8.23. The existence argument will be an easy corollary. A by-product of the existence argument will be a way of testing a polynomial for membership in I.

In the one-variable case any ideal is principal, necessarily of the form $(g(X))$, and the test for membership of a polynomial f in the ideal is to apply the division algorithm, writing $f(X) = q(X)g(X) + r(X)$ with $r = 0$ or $\deg r < \deg g$. Then f is a member of the ideal if and only if $r = 0$. The starting point for the several-variable theory is to do the best we can to generalize the division algorithm to several variables, recognizing that we cannot expect too much because of the complicated ideal structure in several variables.

Proposition 8.20 (generalized division algorithm). Let $(f_1, \dots, f_s)$ be a fixed enumeration of a set of nonzero members of $K[X_1, \dots, X_n]$, and let f be an arbitrary nonzero member of $K[X_1, \dots, X_n]$. Then there exist polynomials $a_1, \dots, a_s$ and r such that

$$f = a_1 f_1 + \cdots + a_s f_s + r,$$

such that $\text{LM}(a_j f_j) \leq \text{LM}(f)$ for all j, and such that no monomial appearing in r with nonzero coefficient is divisible by $\text{LM}(f_j)$ for any j.

REMARK. The proof below will stop short of giving an algorithm, because omitting the details of the algorithm will make the invariant of the construction clearer. To make the proof into an algorithm, one merely needs to be systematic about the choices in the proof. There is no claim of any uniqueness of $a_1, \dots, a_s$ or r in the statement; in fact, Problem 16 at the end of the chapter shows that more than one kind of nonuniqueness is possible. Corollary 8.21 below, however, will show that if the given $f_1, \dots, f_s$ form a Gröbner basis of an ideal I, then r is independent of the enumeration of the Gröbner basis, even without the requirement that $\text{LM}(a_j f_j) \leq \text{LM}(f)$ for all j.

PROOF. We shall do a kind of induction involving decompositions of f of the form

$$f = (a_1 f_1 + \cdots + a_s f_s) + p + r, \tag{$*$}$$

where $a_1, \dots, a_s, p, r$ are polynomials with the properties that

(i) $\text{LM}(p) \leq \text{LM}(f)$,
(ii) $\text{LM}(a_i f_i) \leq \text{LM}(f)$ for all i,
(iii) no monomial M appearing in r with nonzero coefficient has M divisible by any $\text{LM}(f_i)$,

and we shall demonstrate that $\text{LM}(p)$ decreases at every step of the induction as long as $p \neq 0$. Initially we take all $a_i = 0$, $p = f$, and $r = 0$. Then $(*)$ and the three properties hold at the start. Let us describe the inductive step.

If $\text{LT}(f_j)$ divides $\text{LT}(p)$ for some j, then we replace a_j by $a_j + \text{LT}(p)/\text{LT}(f_j)$, we change p to $p - \big(\text{LT}(p)/\text{LT}(f_j)\big) f_j$, and we leave r alone. The equality $(*)$ is maintained, and (iii) continues to hold. Since

$$\begin{aligned} \text{LT}\Big(\big(\text{LT}(p)/\text{LT}(f_j)\big) f_j\Big) &= \text{LT}\big(\text{LT}(p)/\text{LT}(f_j)\big)\,\text{LT}(f_j) \\ &= \big(\text{LT}(p)/\text{LT}(f_j)\big)\,\text{LT}(f_j) = \text{LT}(p), \end{aligned} \tag{$**$}$$

Proposition 8.18 shows that $\text{LM}(p)$ strictly decreases. Consequently (i) continues to hold. By the same kind of computation as for $(**)$,

$$\begin{aligned} \text{LM}\Big(\big(a_j + \text{LT}(p)/\text{LT}(f_j)\big) f_j\Big) &\leq \max\Big(\text{LM}(a_j f_j), \text{LM}\big(\text{LT}(p)/\text{LT}(f_j)\big) f_j\Big) \\ &\leq \max(\text{LM}(f), \text{LM}(p)) = \text{LM}(f), \end{aligned}$$

and therefore (ii) continues to hold. This completes the inductive step if $\text{LT}(f_j)$ divides $\text{LT}(p)$ for some j.

The contrary case is that $\text{LT}(p)$ is divisible by $\text{LT}(f_i)$ for no i. Then we replace p by $p - \text{LT}(p)$, we change r to $r + \text{LT}(p)$, and we leave all a_i alone. The equality $(*)$ is maintained, and (ii) continues to hold. Since $\text{LM}(p) = \text{LM}(\text{LT}(p))$, Proposition 8.18 shows that $\text{LM}(p)$ strictly decreases. Consequently (i) continues to hold. Also, (iii) continues to hold because of the assumption that $\text{LT}(p)$ is divisible by $\text{LT}(f_i)$ for no i. This completes the inductive step if $\text{LT}(p)$ is divisible by $\text{LT}(f_i)$ for no i.

Proposition 8.16 shows that the induction can continue for only finitely many steps. Since it must continue as long as $p \neq 0$, the conclusion is that $p = 0$ after some stage, and then the decomposition of the proposition has been proved. □

Corollary 8.21. If $\{g_1, \dots, g_s\}$ is a Gröbner basis of a nonzero ideal I of $K[X_1, \dots, X_n]$ and if f is any nonzero member of $K[X_1, \dots, X_n]$, then there exist polynomials g and r such that $f = g + r$, g is in I, and no monomial appearing in r with nonzero coefficient is divisible by $\text{LM}(g_j)$ for any j. Moreover, r is uniquely determined by these properties, and g has an expansion $g = \sum_{i=1}^{s} a_i g_i$ with $\text{LM}(a_i g_i) \leq \text{LM}(f)$ for all i.

REMARKS. The uniqueness statement implies in particular that r is independent of the enumeration of the set $\{g_1, \dots, g_s\}$. This corollary will give us some insight into the way a Gröbner basis can resolve cancellation. Shortly we shall introduce specific members of I that have cancellation built into their definition. Being in I, they have expansions with remainder term 0, according to this corollary. Since the remainder is unique, the corollary says that they can be rewritten in terms of the Gröbner basis in a way that eliminates the cancellation.

PROOF. For existence, let $\{g_1, \dots, g_s\}$ be a Gröbner basis of I, and apply Proposition 8.20 to f and the ordered set $(g_1, \dots, g_s)$. Then the existence follows immediately.

For uniqueness, suppose that $f = g_1 + r_1 = g_2 + r_2$. Then $r_1 - r_2 = g_2 - g_1$ exhibits $r_1 - r_2$ as in I. Arguing by contradiction, suppose that $r_1 \neq r_2$. The hypothesis on r_1 and r_2 shows that no monomial with nonzero coefficient in $r_1 - r_2$ is divisible by any $\text{LM}(g_j)$, and in particular $\text{LM}(r_1 - r_2)$ is not divisible by any of the generators of the monomial ideal $\big(\text{LM}(g_1), \dots, \text{LM}(g_s)\big) = \text{LM}(I)$. Since $\text{LM}(r_1 - r_2)$ is a monomial in this ideal, this conclusion contradicts the last conclusion of Lemma 8.17. □

Suppose that $X^\alpha = X_1^{\alpha_1} \cdots X_n^{\alpha_n}$ and $X^\beta = X_1^{\beta_1} \cdots X_n^{\beta_n}$ are two monomials in $K[X_1, \dots, X_n]$. Then we define their **least common multiple** $\text{LCM}(X^\alpha, X^\beta)$ to be

$$\text{LCM}(X^\alpha, X^\beta) = X^\gamma = X_1^{\gamma_1} \cdots X_n^{\gamma_n} \qquad \text{with } \gamma_j = \max(\alpha_j, \beta_j) \text{ for all } j.$$

This notion does not depend on the choice of a monomial ordering. Observe for any two monomials M and N that $\text{LCM}(M, N)/M$ and $\text{LCM}(M, N)/N$ are monomials.

If f_1 and f_2 are nonzero polynomials, then the expression

$$\frac{\mathrm{LCM}\big(\mathrm{LM}(f_1), \mathrm{LM}(f_2)\big)}{\mathrm{LT}(f_1)} f_1 = \frac{\mathrm{LCM}\big(\mathrm{LM}(f_1), \mathrm{LM}(f_2)\big)}{\mathrm{LM}(f_1)} \frac{f_1}{\mathrm{LC}(f_1)}$$

is a polynomial whose leading monomial is $\mathrm{LCM}\big(\mathrm{LM}(f_1), \mathrm{LM}(f_2)\big)$ and whose leading coefficient is 1. We define the **S-polynomial** of f_1 and f_2 to be

$$S(f_1, f_2) = \frac{\mathrm{LCM}\big(\mathrm{LM}(f_1), \mathrm{LM}(f_2)\big)}{\mathrm{LT}(f_1)} f_1 - \frac{\mathrm{LCM}\big(\mathrm{LM}(f_1), \mathrm{LM}(f_2)\big)}{\mathrm{LT}(f_2)} f_2.$$

This is the difference of two polynomials with the same leading monomial $\mathrm{LCM}\big(\mathrm{LM}(f_1), \mathrm{LM}(f_2)\big)$ and with the same leading coefficient 1. Accordingly, Proposition 8.18 shows that

$$\mathrm{LM}(S(f_1, f_2)) < \mathrm{LCM}\big(\mathrm{LM}(f_1), \mathrm{LM}(f_2)\big).$$

The elements $S(f_1, f_2)$ are the elements mentioned in the remarks with Corollary 8.21; the above inequality is a precise formulation of their built-in cancellation.

Lemma 8.22 below says that whenever cancellation of this kind occurs in any sum of products with functions $f_1, \dots, f_s$, then the sum of products can be rewritten in terms of the S-polynomials $S(f_j, f_k)$. In this way the nature of the cancellation has been made more transparent, partly being accounted for by the definitions of the individual polynomials $S(f_j, f_k)$.

Lemma 8.22. Let M and $M_1, \dots, M_s$ be monomials, let $f_1, \dots, f_s$ be nonzero polynomials, and suppose that $M_i \,\mathrm{LM}(f_i) = M$ for all i. If $c_1, \dots, c_s$ are constants such that $\mathrm{LM}\left(\sum_{i=1}^s c_i M_i f_i\right) < M$, then the sum $\sum_{i=1}^s c_i M_i f_i$ can be rewritten in the form

$$\sum_{i=1}^s c_i M_i f_i = \sum_{j<k} \frac{d_{jk} M}{\mathrm{LCM}\big(\mathrm{LM}(f_j), \mathrm{LM}(f_k)\big)} S(f_j, f_k)$$

for suitable constants d_{jk}. In the sum on the right side, each nonzero term has leading monomial $< M$.

PROOF. Let us write $L_{ij} = \mathrm{LCM}\big(\mathrm{LM}(f_i), \mathrm{LM}(f_j)\big)$ for $i \neq j$. We may assume that all the c_i are nonzero, and we proceed by induction on s. There is nothing to prove for $s = 1$. The key step is $s = 2$, for which we are given that the M term of $c_1 M_1 f_1 + c_2 M_2 f_2$ is 0, i.e., that

$$c_1 \,\mathrm{LC}(f_1) + c_2 \,\mathrm{LC}(f_2) = 0. \qquad (*)$$

Substituting for $\mathrm{LC}(f_2)$ from $(*)$ gives

$$\begin{aligned} ML_{12}^{-1}S(f_1, f_2) &= Mf_1/\mathrm{LT}(f_1) - Mf_2/\mathrm{LT}(f_2) \\ &= M_1f_1/\mathrm{LC}(f_1) - M_2f_2/\mathrm{LC}(f_2) \\ &= c_1^{-1}\,\mathrm{LC}(f_1)^{-1}(c_1M_1f_1 + c_2M_2f_2), \end{aligned}$$

and this proves the displayed formula of the lemma with $d_{12} = c_1\,\mathrm{LC}(f_1)$.

Assume the result for $s-1 \geq 2$. We are given that $\sum_{i=1}^{s} c_i\,\mathrm{LC}(f_i) = 0$, which we break into two parts as

$$\begin{gathered} c_1\,\mathrm{LC}(f_1) - \tfrac{c_1\,\mathrm{LC}(f_1)}{\mathrm{LC}(f_2)}\,\mathrm{LC}(f_2) = 0, \\ \Big(c_2 + \tfrac{c_1\,\mathrm{LC}(f_1)}{\mathrm{LC}(f_2)}\Big)\,\mathrm{LC}(f_2) + \sum_{i=3}^{s} c_i\,\mathrm{LC}(f_i) = 0. \end{gathered}$$

The inductive hypothesis gives

$$\begin{gathered} c_1M_1f_1 - \tfrac{c_1\,\mathrm{LC}(f_1)}{\mathrm{LC}(f_2)}\,M_2f_2 = d_{12}ML_{12}^{-1}S(f_1, f_2), \\ \Big(c_2 + \tfrac{c_1\,\mathrm{LC}(f_1)}{\mathrm{LC}(f_2)}\Big)M_2f_2 + \sum_{i=3}^{s} c_iM_if_i = \sum_{2\leq j<k} d_{jk}ML_{jk}^{-1}S(f_j, f_k). \end{gathered}$$

Adding these two formulas, we obtain the displayed formula of the lemma for the case s, and the induction is complete. □

Theorem 8.23. Let $\{g_1, \ldots, g_s\}$ be a set of generators of a nonzero ideal I of $K[X_1, \ldots, X_n]$, and assume that $g_i \neq 0$ for all i. Then the following conditions on $\{g_1, \ldots, g_s\}$ are equivalent:

(a) $\{g_1, \ldots, g_s\}$ is a Gröbner basis of I,

(b) for each pair (g_j, g_k) with $S(g_j, g_k) \neq 0$, every expansion of $S(g_j, g_k)$ as $S(g_j, g_k) = \sum_{i=1}^{s} a_{ijk}g_i + r$ with the two properties that

(i) $\mathrm{LM}(a_{ijk}g_i) \leq \mathrm{LM}(S(g_j, g_k))$ and

(ii) no monomial appearing in r with nonzero coefficient is divisible by $\mathrm{LM}(g_j)$ for any j

has $r = 0$,

(c) for each pair (g_j, g_k) with $S(g_j, g_k) \neq 0$, there is an expansion of the form $S(g_j, g_k) = \sum_{i=1}^{s} a_{ijk}g_i$ with $\mathrm{LM}(a_{ijk}g_i) \leq \mathrm{LM}(S(g_j, g_k))$.

REMARKS. Because of the equivalence of (b) and (c), the generalized division algorithm (Proposition 8.20) gives us a procedure for testing whether these conditions are satisfied by $\{g_1, \ldots, g_s\}$. Namely we follow through the steps in the proof of Proposition 8.20 in whatever fashion we please for each nonzero

$S(g_j, g_k)$. If we get remainder $r = 0$ for each pair (j, k), then the conditions are satisfied. If we get a nonzero remainder r for some pair, then the conditions are not satisfied. In view of the equivalence of (a) with these conditions, we have an effective (though somewhat tedious) way of checking whether $\{g_1, \dots, g_s\}$ is a Gröbner basis.

PROOF. We prove that (a) implies (b) and that (c) implies (a). Since (b) certainly implies (c), the proof will be complete.

Let (a) hold, i.e., let $\{g_1, \dots, g_s\}$ be a Gröbner basis. If $S(g_j, g_k) \neq 0$, then $S(g_j, g_k)$ is a nonzero member of I because each g_i lies in I, and $S(g_j, g_k)$ consequently has an expansion as $\sum_{i=1}^{s} a_i g_i + r$ with $r = 0$. By Corollary 8.21 it has a possibly different expansion with $r = 0$ and with $\mathrm{LM}(a_i g_i) \leq \mathrm{LM}(S(g_j, g_k))$ for each i. On the other hand, in any expansion of $S(g_j, g_k)$ as $\sum_{i=1}^{s} a_i g_i + r$ such that (ii) holds, whether or not $\mathrm{LM}(a_i g_i) \leq \mathrm{LM}(S(g_j, g_k))$, r must be 0 by Corollary 8.21. This proves (b).

To prove that (c) implies (a), we argue by contradiction. Among all expansions of members of I as $\sum_{i=1}^{s} b_i g_i$ such that $\mathrm{LT}\left(\sum_{i=1}^{s} b_i g_i\right)$ is not in the ideal $\left(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_s)\right)$, choose one for which

$$M = \max_{1 \leq i \leq s} \mathrm{LM}(b_i g_i)$$

is as small as possible; this choice exists by Proposition 8.16. For this choice, let

$$f = \sum_{i=1}^{s} b_i g_i. \tag{$*$}$$

Define $M_i = \mathrm{LM}(b_i)$ for each i with $b_i \neq 0$. If i_0 is an index with $M = \mathrm{LM}(b_{i_0} g_{i_0})$, then $M = M_{i_0} \mathrm{LM}(g_{i_0})$ by Proposition 8.18, and hence M lies in $\left(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_s)\right)$. Since $\mathrm{LT}\left(\sum_{i=1}^{s} b_i g_i\right)$ is not in $\left(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_s)\right)$, it follows that $\mathrm{LT}\left(\sum_i b_i g_i\right) < M$. Within the set $\{1, \dots, s\}$, define a subset E to consist of those i for which $M_i \, \mathrm{LM}(g_i) = M$. This set contains i_0, and it has the property that all i not in E have $\mathrm{LM}(b_i g_i) < M$. We regroup f as

$$f = \sum_{i \in E} b_i g_i + \sum_{i \notin E} b_i g_i = \sum_{i \in E} \mathrm{LC}(b_i) M_i g_i + \sum_{i \in E} (b_i - \mathrm{LT}(b_i)) g_i + \sum_{i \notin E} b_i g_i.$$

Every term in the second and third sums on the right side has leading monomial $< M$, and so does f. Therefore $\mathrm{LM}\left(\sum_{i \in E} \mathrm{LC}(b_i) M_i g_i\right) < M$. It follows that the expression $\sum_{i \in E} \mathrm{LC}(b_i) M_i g_i$ is of the form considered in Lemma 8.22 with $c_i = \mathrm{LC}(b_i)$ for $i \in E$ (and $c_i = 0$ for $i \notin E$). The lemma tells us that

$$\sum_{i \in E} \mathrm{LC}(b_i) M_i g_i = \sum_{j,k} d_{jk} (M / L_{jk}) S(g_j, g_k)$$

for suitable scalars d_{jk}, where $L_{jk} = \text{LCM}\big(\text{LM}(g_j), \text{LM}(g_k)\big)$.

Now we apply the hypothesis (c), expanding each $S(g_j, g_k)$ in some way as $S(g_j, g_k) = \sum_{i=1}^{s} a_{ijk} g_i$ with the a_{ijk} equal to polynomials such that

$$\text{LM}(a_{ijk} g_i) \le \text{LM}(S(g_j, g_k)). \qquad (**)$$

Substituting for $S(g_j, g_k)$, we obtain

$$f = \sum_{i,j,k} d_{jk}(M/L_{jk}) a_{ijk} g_i + \sum_{i \in E} (b_i - \text{LT}(b_i)) g_i + \sum_{i \notin E} b_i g_i. \qquad (\dagger)$$

We know that every term in the second and third sums on the right side of (†) has leading monomial $< M$, and we shall estimate the leading monomial of each term in the first sum. Multiplying the inequality

$$\text{LM}(S(g_j, g_k)) < \text{LCM}\big(\text{LM}(g_j), \text{LM}(g_k)\big) = L_{jk}$$

by the monomial M/L_{jk} yields

$$(M/L_{jk})\, \text{LM}(S(g_j, g_k)) < M \qquad (\dagger\dagger)$$

for every pair (j, k). Combining $(**)$ and $(\dagger\dagger)$ gives

$$\text{LM}\left((M/L_{jk}) a_{ijk} g_i\right) = (M/L_{jk})\, \text{LM}(a_{ijk} g_i) \le (M/L_{jk})\, \text{LM}(S(g_j, g_k)) < M.$$

Since each d_{jk} is a scalar, every term in the first sum on the right side of (†) has leading monomial $< M$. Thus (†) is an expansion of a member of I that contradicts the minimality of $\max_i \text{LM}(b_i g_i)$ in the expansion $(*)$. From this contradiction we conclude that (a) holds. □

EXAMPLE OF A VERIFICATION THAT A SET IS A GRÖBNER BASIS. This example continues Example 2 of "Examples with lexicographic ordering" in the previous section. A nonzero ideal I is generated by members of $K[X_1, \ldots, X_n]_1$ of the form $(L_1, \ldots, L_s)$, where each L_j is a linear combination of $X_1, \ldots, X_n$. After initial manipulations we assume that the matrix of coefficients of $L_1, \ldots, L_s$ is in reduced row-echelon form. The assertion is that $\{L_1, \ldots, L_s\}$ is then a Gröbner basis of I. To prove this, we write $L_j = X_{n_j} + l_j$, where X_{n_j} is the associated corner variable and l_j is a linear combination of $X_{n_j+1}, \ldots, X_n$ such that the coefficient of each corner variable is 0. If $j < k$, then

$$S(L_j, L_k) = -l_k X_{n_j} + l_j X_{n_k} = -l_k(X_{n_j} + l_j) + l_j(X_{n_k} + l_k) = -l_k L_j + l_j L_k.$$

The second term on the right side contains no variable $X_1, \ldots, X_{n_j}$, but the first term on the right side contains X_{n_j}. Therefore, relative to the lexicographic ordering, we have $\text{LM}\left(S(L_j, L_k)\right) = \text{LM}(-l_k L_j) = \text{LM}(l_k) X_{n_j}$. Consequently $\text{LM}(l_j L_k) \le \text{LM}\left(S(L_j, L_k)\right)$ (and actually strict inequality must hold). Thus the displayed formula shows that $S(L_j, L_k) = a_1 L_j + a_2 L_k$ in the form demanded by (c) of Theorem 8.23. Since (c) implies (a) in the theorem, $\{L_1, \ldots, L_s\}$ is a Gröbner basis of I.

Corollary 8.24 (Buchberger's algorithm).[15] Each nonzero ideal in the polynomial ring $K[X_1, \dots, X_n]$ has a Gröbner basis. Such a basis can be obtained by the following procedure: Start from any set $\{f_1, \dots, f_t\}$ of nonzero generators, apply the generalized division algorithm in some fashion to each $S(f_j, f_k)$ and to the generating set $\{f_1, \dots, f_t\}$, and adjoin to the set of generators any nonzero remainders obtained from this process. Iterate this process for enlarging a set $\{f_1', \dots, f_{t'}'\}$ of generators as long as a nonzero remainder is obtained for some $S(f_j', f_k')$. This process must terminate at some point with all remainders equal to 0, and the resulting generating set is a Gröbner basis.

PROOF. At the stage of the iteration that works with the set $\{f_1', \dots, f_{t'}'\}$ of generators, any nonzero remainder r that arises has the property that no monomial occurring in r is divisible by any $\mathrm{LM}(f_j')$. By Lemma 8.17, $\mathrm{LT}(r)$ is not a member of $\big(\mathrm{LT}(f_1'), \dots, \mathrm{LT}(f_t')\big)$. However, at the next stage when r has been designated as one of the generators of I, $\mathrm{LT}(r)$ has become one of the generators of this ideal. Therefore the ideal $\big(\mathrm{LT}(f_1'), \dots, \mathrm{LT}(f_t')\big)$ strictly increases as we pass from one stage to the next. Since $K[X_1, \dots, X_n]$ is Noetherian, its ideals satisfy the ascending chain condition, and this chain of ideals must stabilize. Consequently all the remainders must be 0 at some point, and then Theorem 8.23 shows that the set of generators is a Gröbner basis. □

EXAMPLE OF THE COMPUTATION OF A GRÖBNER BASIS. We return to Example 3 of "Examples with lexicographic ordering" in the previous section. In $K[X, Y]$, we let $f_1(X, Y) = X^2 + 2XY^2$ and $f_2(X, Y) = XY + 2Y^3 - 1$, and we define $I = (f_1, f_2)$. We seek a Gröbner basis of I, using the lexicographic ordering. Direct computation gives $S(f_1, f_2) = Y(X^2 + 2XY^2) - X(XY + 2Y^3 - 1) = X$. Since X is not divisible by $\mathrm{LM}(f_1)$ or by $\mathrm{LM}(f_2)$, $S(f_1, f_2) = 0f_1 + 0f_2 + X$ is an expansion of $S(f_1, f_2)$ as in Theorem 8.23c with $r = X$. The procedure of Corollary 8.24 says to adjoin $f_3 = X$ to the generating set and test again. Direct computation gives $S(f_1, f_3) = 1(X^2 + 2XY^2) - X \cdot X = 2XY$, and $S(f_1, f_3) = 0f_1 + 0F_2 + (2Y)f_3 + 0$ is an expansion of $S(f_1, f_3)$ as in (c), since $\mathrm{LM}(2Yf_3) \leq \mathrm{LM}\big(S(f_1, f_3)\big)$. Thus $S(f_1, f_3)$ gives us a 0 remainder, hence nothing new to process. In addition, we have $S(f_2, f_3) = 1(XY + 2Y^3 - 1) - Y \cdot X = 2Y^3 - 1$. No term of this is divisible by any of the leading monomials of f_1, f_2, f_3, namely X^2, XY, X. Hence $2Y^3 - 1$ is a nonzero remainder.[16] Therefore we are to adjoin $f_4 = 2Y^3 - 1$ to our set. Computation gives $S(f_1, f_4) = 2XY^4 + X^2 = (2Y^4 + X)f_3$, $S(f_2, f_4) = 2Y^5 - Y^2 + \frac{1}{2}X = \frac{1}{2}f_3 + Y^2 f_4$,

[15] Computer programs typically use an improved version of this algorithm to compute Gröbner bases.

[16] It was not a bad choice of decomposition that led to a nonzero remainder when some other decomposition might have given us 0; the equivalence of (b) and (c) in Theorem 8.23 assures us of that fact.

and $S(f_3, f_4) = \frac{1}{2}X = \frac{1}{2}f_3$. In every case each term has leading monomial at most the leading monomial of the S-polynomial. Hence all remainders are 0, and Corollary 8.24 says that $\{f_1, f_2, f_3, f_4\}$ is a Gröbner basis of I.

Corollary 8.25 (solution of the ideal-membership problem). If I is a nonzero ideal in $K[X_1, \dots, X_n]$ and f is a polynomial, then a procedure for deciding whether f lies in I is as follows: introduce a monomial ordering, construct a Gröbner basis $\{g_1, \dots, g_s\}$ of I by means of Corollary 8.24, and apply the generalized division algorithm to write $f = \sum_{i=1}^{s} a_i g_i + r$ for polynomials $a_1, \dots, a_r, r$ such that no monomial appearing in r with nonzero coefficient is divisible by $\mathrm{LM}(g_j)$ for any j. Then f lies in I if and only if $r = 0$.

PROOF. Corollary 8.24 produces the Gröbner basis, and Corollary 8.21 affirms that this procedure decides whether f lies in I. □

Corollary 8.26 (solution of the proper-ideal problem). If I is a nonzero ideal in $K[X_1, \dots, X_n]$, then a procedure for deciding whether $I = K[X_1, \dots, X_n]$ is to compute a Gröbner basis for I and to see whether one of its members is a nonzero scalar c.

PROOF. If I has a nonzero scalar as one of its generators, then 1 lies in I, and hence I certainly equals $K[X_1, \dots, X_n]$. Conversely if I is given, then Corollary 8.24 produces a Gröbner basis $\{g_1, \dots, g_s\}$. Since $\mathrm{LT}(1) = 1$ and since $\mathrm{LT}(I) = \big(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_s)\big)$, the monomial 1 must lie in $\big(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_s)\big)$. Since 1 is a monomial, Lemma 8.17 shows that it must be divisible by $\mathrm{LM}(g_j)$ for some j. Therefore $\mathrm{LM}(g_j) = 1$. Since 1 is the smallest monomial in any monomial ordering, it is the only monomial appearing with a nonzero coefficient in g_j. Therefore g_j is a nonzero scalar. □

In many applications of Gröbner bases, there is some flexibility in what monomial ordering to impose in obtaining the Gröbner basis. In Corollaries 8.25 and 8.26, for example, absolutely any monomial ordering works fine. The actual calculation of Gröbner bases is often computationally demanding, and thus it is worthwhile to use such a basis that takes relatively little time to compute. According to computer scientists,[17] Gröbner bases are the most widely useful when computed relative to the lexicographic ordering, but they are then also the most time-consuming to compute. The monomial orderings that make the computation of Gröbner bases proceed quickly tend to be ones that first bound

[17] The Web essay "Representation and monomial orders," `http://magma.usyd.edu/au/magma/handbook/1177`, within the documentation of the Magma computer algebra system at the University of Sydney contains a discussion of various monomial orders and their uses and advantages.

the total degree in one or two steps. One of the reasons that this kind of monomial ordering works so efficiently is that once the total degree is bounded, there are only finitely many monomials less than any given monomial M.

9. Uniqueness of Reduced Gröbner Bases

In this section, K continues to denote a field, and we work with a fixed monomial ordering on $K[X_1, \dots, X_n]$. Ideals in $K[X_1, \dots, X_n]$ will always be specified by giving finite sets of generators. Our objective in this section is to show how any Gröbner basis can be "reduced" and that a "reduced" Gröbner basis for an ideal is unique. A by-product of the uniqueness argument will be a way of testing two ideals for equality.

Any finite set of generators of I that contains a Gröbner basis is again a Gröbner basis. Thus a constructed Gröbner basis will often be unnecessarily large. One simple kind of redundance is addressed by Lemma 8.27 below.

Lemma 8.27. If $\{g_1, \dots, g_s\}$ is a Gröbner basis for a nonzero ideal I in $K[X_1, \dots, X_n]$ and if $\text{LM}(g_1)$ lies in the ideal $\big(\text{LT}(g_2), \dots, \text{LT}(g_s)\big)$, then $\{g_2, \dots, g_s\}$ is a Gröbner basis of I.

REMARK. Lemma 8.17 shows how to check whether $\text{LM}(g_1)$ lies in the ideal $\big(\text{LT}(g_2), \dots, \text{LT}(g_s)\big)$; all we have to do is see whether some $\text{LM}(g_j)$ for $j \geq 1$ divides $\text{LM}(g_1)$.

PROOF. By hypothesis, $\big(\text{LT}(g_2), \dots, \text{LT}(g_s)\big) = \big(\text{LT}(g_1), \dots, \text{LT}(g_s)\big) = \text{LT}(I)$. Therefore $\{g_2, \dots, g_s\}$ is a Gröbner basis of I. (Recall that the definition of Gröbner basis does not assume that the set generates the ideal; Proposition 8.19 deduces that it generates.) □

A Gröbner basis $\{g_1, \dots, g_s\}$ of a nonzero ideal I is said to be **minimal** if $\text{LC}(g_j) = 1$ for all j and if no $\text{LM}(g_i)$ is divisible by $\text{LM}(g_j)$ for some $j \neq i$. Lemma 8.27 shows that in trying to transform a Gröbner basis into a form for which a uniqueness result will apply, there is no loss of generality in assuming that the given Gröbner basis is minimal.

EXAMPLE. As in the example following Corollary 8.24, let I be the ideal in $K[X, Y]$ given by $I = (f_1, f_2)$ with $f_1(X, Y) = X^2 + 2XY^2$ and $f_2(X, Y) = XY + 2Y^3 - 1$. Then we saw that $\{f_1, f_2, f_3, f_4\}$ is a Gröbner basis of I in the lexicographic ordering, where $f_3(X, Y) = X$ and $f_4(X, Y) = 2Y^3 - 1$. The leading monomials are $\text{LM}(f_1) = X^2$, $\text{LM}(f_2) = XY$, $\text{LM}(f_3) = X$, and $\text{LM}(f_4) = Y^3$. The first two are divisible by the third. Therefore $\{X, Y^3 - \frac{1}{2}\}$ is the corresponding minimal Gröbner basis.

Unfortunately an ideal can have more than one minimal Gröbner basis, as is shown in Problem 17 at the end of the chapter. A Gröbner basis $\{g_1, \dots, g_s\}$ of an ideal I is said to be **reduced** if it is minimal and if for each i, no monomial appearing in g_i with nonzero coefficient is divisible by $\mathrm{LM}(g_j)$ for some $j \neq i$.

Theorem 8.28 (uniqueness of reduced Gröbner basis). If I is a nonzero ideal in $K[X_1, \dots, X_n]$, then I has a unique reduced Gröbner basis, and this can be obtained algorithmically starting from any minimal Gröbner basis.

PROOF OF UNIQUENESS. Let $\{g_1, \dots, g_s\}$ be any Gröbner basis. Since $\mathrm{LT}(I) = \big(\mathrm{LT}(g_1), \dots, \mathrm{LT}(g_s)\big)$, Lemma 8.17 shows that any $\mathrm{LM}(f)$ for $f \in I$ is divisible by $\mathrm{LM}(g_j)$ for some j. If $\{h_1, \dots, h_t\}$ is a second Gröbner basis, then this argument shows that each $\mathrm{LM}(h_i)$ is divisible by some $\mathrm{LM}(g_j)$. Turned around, the argument shows that $\mathrm{LM}(g_j)$ is divisible by some $\mathrm{LM}(h_k)$. Since $\{h_1, \dots, h_t\}$ is assumed minimal, $\mathrm{LM}(h_k)$ cannot be divisible by $\mathrm{LM}(h_i)$ if $i \neq k$. Thus $\mathrm{LM}(h_i) = \mathrm{LM}(h_k)$, and these equal $\mathrm{LM}(g_j)$. Then it follows that $s = t$ and that we may enumerate any two minimal Gröbner bases in such a way that the leading monomial of the i^{th} member of each basis is the same for each i with $1 \leq i \leq s$.

With this normalization in place, let us show that $g_i = h_i$. To do so, we expand $g_i - h_i$ as $g_i - h_i = \sum_{j=1}^{s} a_j h_j$ with $\mathrm{LM}(g_i - h_i) = \max_j \mathrm{LM}(a_j h_j)$ in accordance with (b) of Theorem 8.23. Choose k such that the maximum on the right side is attained at k, i.e., such that

$$\mathrm{LM}(a_k)\,\mathrm{LM}(h_k) = \mathrm{LM}(g_i - h_i). \tag{$*$}$$

Arguing by contradiction, suppose that the right side of $(*)$ is nonzero. Then it must be a monomial occurring in either g_i or h_i. Since the two Gröbner bases are reduced, no monomial occurring in g_i is divisible by $\mathrm{LM}(g_k) = \mathrm{LM}(h_k)$ if $k \neq i$, and similarly for monomials occurring in h_i. We conclude that $k = i$ and that $\mathrm{LM}(h_i) = \mathrm{LM}(g_i - h_i)$. But this is impossible by Proposition 8.18 if $g_i - h_i \neq 0$, since $\mathrm{LM}(g_i) = \mathrm{LM}(h_i)$ and $\mathrm{LC}(g_i) = \mathrm{LC}(h_i) = 1$. Therefore the right side of $(*)$ is 0, and $g_i = h_i$. □

PROOF OF EXISTENCE. Let $\{g_1, \dots, g_s\}$ be a minimal Gröbner basis of I. As was shown in the proof of uniqueness, the leading monomials $\mathrm{LM}(g_1), \dots, \mathrm{LM}(g_s)$ are independent of the choice of the actual minimal basis. Looking at the definition of "reduced," we see therefore that the property of being reduced is a property of each member g_i of the basis separately. That is, it is meaningful to say that g_i is reduced if no monomial appearing in g_i with nonzero coefficient is divisible by $\mathrm{LM}(g_j)$ for some $j \neq i$. We shall show how to replace g_i by an element g_i' with the same leading monomial in such a way that the new set is still a Gröbner basis and g_i' is reduced, and then the proof will be complete. There is no loss of generality in taking $i = 1$.

Applying the generalized division algorithm (Proposition 8.20), we write

$$g_1 = \sum_{j=2}^{s} a_j g_j + r \tag{$**$}$$

in such a way that

$$\mathrm{LM}(g_1) = \max_{2 \le j \le s} \mathrm{LM}(a_j g_j) \tag{$\dagger$}$$

and that no monomial appearing in r with nonzero coefficient is divisible by $\mathrm{LM}(g_j)$ for any $j \ge 2$. If we define g_1' to be this element r, then the element g_1' is reduced in the above sense, and the only question is whether $\{g_1', g_2, \dots, g_s\}$ is a Gröbner basis. Since $\{g_1, \dots, g_s\}$ is minimal, $\mathrm{LM}(g_1)$ is not divisible by any $\mathrm{LM}(g_j)$ for $j \ge 2$. Consequently $\mathrm{LM}(g_1)$ appears with nonzero coefficient on the left side of $(**)$, and it does not appear in any of the terms $a_j g_j$ with nonzero coefficient on the right side. Consequently it appears in $r = g_1'$, and $\mathrm{LM}(g_1) \le \mathrm{LM}(g_1')$. On the other hand, the equality $(\dagger)$ implies that $\mathrm{LM}(g_1') \le \mathrm{LM}(g_1)$. Therefore $\mathrm{LM}(g_1) = \mathrm{LM}(g_1')$, and $\mathrm{LT}(I) = \big(\mathrm{LT}(g_1), \mathrm{LT}(g_2) \dots, \mathrm{LT}(g_s)\big) = \big(\mathrm{LT}(g_1'), \mathrm{LT}(g_2) \dots, \mathrm{LT}(g_s)\big)$. Consequently $\{g_1', g_2, \dots, g_s\}$ is a Gröbner basis by definition. □

Corollary 8.29 (solution of the ideal-equality problem). Let I and J be two nonzero ideals in $K[X_1, \dots, X_n]$ specified in terms of finite sets of generators. Then $I = J$ if and only if the reduced Gröbner bases of I and J relative to a single monomial ordering are the same.

REMARK. As with the solution of problems listed in Corollaries 8.25 and 8.26, the desired end is independent of the monomial ordering, and in practice one might just as well start from a monomial ordering for which the computation of Gröbner bases is relatively easy.

PROOF. This result is immediate from Corollary 8.24 (constructive existence of Gröbner bases) and Theorem 8.28. □

10. Simultaneous Systems of Polynomial Equations

In this section we combine our techniques concerning the resultant and Gröbner bases to attack the original problem discussed in Section 1, that of solving systems of simultaneous polynomial equations in several variables. Our interest ultimately will be in the case that the underlying field is algebraically closed.

Corollary 8.26 and the Nullstellensatz already combine to give a criterion for such a system to have no solutions: We regard the system as the zero locus of an ideal, and we calculate a Gröbner basis for the ideal. Then the system has no solutions if and only if the Gröbner basis contains a constant polynomial, i.e., if and only if the reduced Gröbner basis is $\{1\}$.

Let us now consider the problem of finding the solutions when solutions exist. We begin with the case of two equations in two unknowns over the field $\mathbb{C}$, recalling what we know from the theory of the resultant. Consider the system

$$\begin{aligned} X^2Y + Y^2 &= 5, \\ XY &= 2. \end{aligned}$$

Set $f(X, Y) = X^2Y + Y^2 - 5$ and $g(X, Y) = XY - 2$. To find points (x, y) with $f(x, y) = g(x, y) = 0$, using the style of Sections 1–3, we compute the resultant of f and g in the X variable, say, and obtain the polynomial $Y^4 - 5Y^2 + 4Y$. Setting this equal to 0 gives us $y = 0$, $y = 1$, and $y = \frac{1}{2}(-1 \pm \sqrt{17})$. We can then substitute each such y into $x^2y + y^2 = 5$ and get candidates (x, y). Doing so for $y = 0$ gives us no candidates, and doing so for each of the other three values of y gives us two values of x, differing only in a sign. So we get six pairs (x, y). However, only three of these satisfy the second given equation, $xy = 2$, one for each nonzero value of y. Thus the resultant gives us a handle on the problem of finding solutions, but it has two shortcomings: it produced a value of y yielding no solution pairs (x, y), and it produced extraneous x values.

To find points (x, y) with $f(x, y) = g(x, y) = 0$, using the style of Sections 7–10, we consider (f, g) as an ideal in $\mathbb{C}[X, Y]$, and we are interested in the locus of common zeros $V_{\mathbb{C}}((f, g))$ of the ideal. We start by finding a reduced Gröbner basis with respect to a suitable ordering. The usual lexicographic ordering will do fine here, and the result is $\{X + \frac{1}{2}Y^2 - \frac{5}{2}, Y^3 - 5Y + 4\}$. By what may seem to be good fortune, the second element depends on Y alone, and the roots are $y = 1$ and $y = \frac{1}{2}(-1 \pm \sqrt{17})$. If we substitute these values into the equation $x + \frac{1}{2}y^2 - \frac{5}{2} = 0$, we get one value of x for each y. We can solve because the coefficient 1 of x is nonzero for each y in question. No pair (x, y) that we obtain is superfluous because the locus of common zeros of f and g is identical with the locus of common zeros of the members of the Gröbner basis.

This approach raises several questions about a possible generalization:

(i) Under what conditions can we expect that a Gröbner basis for an ideal I in $K[X, Y]$ will contain a member that depends just on Y?

(ii) If the Gröbner basis contains no element that depends just on Y, then what can we expect?

(iii) If we are able to solve for values of y, under what conditions can we use the remaining member(s) of the Gröbner basis to solve for x?

Part of the answer to (i) is contained in the Elimination Theorem proved as Theorem 8.30 below. This theorem says for the lexicographic ordering that the members of a Gröbner basis that depend just on Y generate $I \cap K[Y]$; in fact, they form a Gröbner basis of this ideal of $K[Y]$. For the case that $I = (f, g)$, the resultant is a member of $I \cap K[Y]$. Thus a nonzero resultant ensures that some member of the Gröbner basis will depend just on Y; on the other hand, $I \cap K[Y]$ has to be a principal ideal in $K[Y]$, and any Gröbner basis of that principal ideal has to contain the ideal's generator (up to a scalar factor). By contrast, a zero resultant leads us to question (ii) because it says, by Theorem 8.1, that f and g have a common factor $h(X, Y)$ of positive degree in X as long as both f and g have positive degree in X. The largest power of X in h has as coefficient a polynomial in Y that has only finitely many roots, and if K is algebraically closed, then every y unequal to one of these roots will produce an x such that $h(x, y) = 0$ and therefore such that $f(x, y) = g(x, y) = 0$. In other words, except in degenerate cases a zero resultant implies that there cannot be a member of the Gröbner basis that depends just on Y. Finally the answer to (iii) lies deeper and is contained in the Extension Theorem, which is proved as Theorem 8.31 below.

Let I be a nonzero ideal in $K[X_1, \dots, X_n]$, K being any field for now. If $0 \leq k \leq n - 1$, then the k^{th} **elimination ideal** of I is the ideal $I \cap K[X_{k+1}, \dots, X_n]$ in $K[X_{k+1}, \dots, X_n]$. A monomial ordering on $K[X_1, \dots, X_n]$ will be said to be of k-**elimination type** if any monomial containing any of $X_1, \dots, X_k$ to a positive power is greater than any monomial in $X_{k+1}, \dots, X_n$ alone. The usual lexicographic ordering is of k-elimination type for every k. An example of a monomial ordering of k-elimination type that is of great interest in applications is the one of Bayer–Stillman described in Example 4 of monomial orderings in Section 7.

Theorem 8.30 (Elimination Theorem). Let K be any field, let I be a nonzero ideal in $K[X_1, \dots, X_n]$, let $0 \leq k \leq n$, and fix a monomial ordering of k-elimination type. If $\{g_1, \dots, g_s\}$ is a Gröbner basis of I, then the subset of members of $\{g_1, \dots, g_s\}$ depending only on $X_{k+1}, \dots, X_n$ is a Gröbner basis of the k^{th} elimination ideal $J = I \cap K[X_{k+1}, \dots, X_n]$.

PROOF. Relabeling the members of $\{g_1, \dots, g_s\}$, we may assume that the g_j's lying in J are $g_1, \dots, g_t$. The first step is to show that $J = (g_1, \dots, g_t)$. If $f \in J$ is given, we apply the generalized division algorithm (Proposition 8.20) and write $f = \sum_{i=1}^{s} a_i g_i + r$ with $\text{LM}(a_i g_i) \leq \text{LM}(f)$ for all i and with no monomial appearing in r with nonzero coefficient divisible by $\text{LM}(g_j)$ for any j. Corollary 8.21 shows that $r = 0$. If $a_i \neq 0$ and i is not $\leq t$, then $\text{LM}(a_i g_i)$ involves at least one of $X_1, \dots, X_k$, and the definition of monomial ordering of k-elimination type implies that $\text{LM}(a_i f_i) > \text{LM}(f)$. It follows that $a_i = 0$ for $i > t$, and thus $J = (g_1, \dots, g_t)$.

To see that $\{g_1, \dots, g_t\}$ is a Gröbner basis of J, we apply Theorem 8.23. We are to show for each pair (g_j, g_k) with $S(g_j, g_k) \neq 0$ and $\{j, k\} \subseteq \{1, \dots, t\}$ that there is an expansion $S(g_j, g_k) = \sum_{i=1}^t a_i g_i$ with $\text{LM}(a_i g_i) \leq \text{LM}\big(S(g_j, g_k)\big)$. In view of the argument with f in the previous paragraph, it is enough to show that $S(g_j, g_k)$ lies in J. The formula is

$$S(g_j, g_k) = \frac{\text{LCM}\big(\text{LM}(g_j), \text{LM}(g_k)\big)}{\text{LT}(g_j)} g_j - \frac{\text{LCM}\big(\text{LM}(g_k), \text{LM}(g_k)\big)}{\text{LT}(g_k)} g_k.$$

The coefficient fractions are members of $K[X_{k+1}, \dots, X_n]$, since the monomial ordering is of k-elimination type, and thus $S(g_j, g_k)$ is indeed in J. □

EXAMPLE. Formula for discriminant of a polynomial in one variable. This example is one that we have addressed before by specialized methods. We include it anyway because the use of Gröbner bases allows one to solve many similar problems that the specialized methods do not address. By way of illustration, let $(X - r)(X - s)(X - t)$ be a cubic polynomial. The discriminant is $D = (r - s)^2(s - t)^2(r - t)^2$. This is a polynomial that is symmetric in r, s, t, and the general theory of symmetric polynomials (in the problems for Chapter VIII in *Basic Algebra*) shows that it has to be a polynomial in the elementary symmetric polynomials $a = r + s + t$, $b = rs + rt + st$, $c = rst$. We seek a formula for D in terms of a, b, c. We form the ideal I in $K[r, s, t, D, a, b, c]$ given by

$$I = \big(D - (r - s)^2(s - t)^2(r - t)^2, a - (r + s + t), b - (rs + rt + st), c - rst\big).$$

With the variables enumerated as r, s, t, D, a, b, c, we use any monomial ordering of 4-elimination type, the lexicographic ordering for example, and form the reduced Gröbner basis of I. Calculation best done with the aid of a computer gives $D - a^2b^2 + 4b^3 + 4a^3c - 18abc + 27c^2$ and three other members of I that involve r, s, or t. Theorem 8.30 shows that the 4th elimination ideal is principal with generator $D - a^2b^2 + 4b^3 + 4a^3c - 18abc + 27c^2$. Thus the desired formula is $D = a^2b^2 - 4b^3 - 4a^3c + 18abc - 27c^2$.

Let us come to the Extension Theorem. The statement and proof of this theorem do not make use of Gröbner bases, but they do refer to the k^{th} elimination ideal, which is identified explicitly in Theorem 8.30 with the aid of a Gröbner basis. The intention is that the theorem be applied inductively in any application, taking into account one additional variable at each step of an induction.

Theorem 8.31 (Extension Theorem). Let K be an algebraically closed field, let $I = (f_1, \dots, f_s)$ be an ideal in $K[X_1, \dots, X_n]$, and let J be the first elimination ideal of I in $K[X_2, \dots, X_n]$. For each f_i, expand f_i in powers of X_1 as

$$f_i(X_1, \dots, X_n) = g_i(X_2, \dots, X_n)X_1^{l_1} + (\text{lower powers of } X_1)$$

with g_i in $K[X_2, \dots, X_n]$ and g_i nonzero unless $f_i = 0$. Suppose that $(c_2, \dots, c_n)$ lies in the zero locus $V_K(J) \subseteq K^{n-1}$. If $g_i(c_2, \dots, c_n) \neq 0$ for some i, then there exists c_1 in K such that $(c_1, \dots, c_n)$ is in the zero locus $V_K(I) \subseteq K^n$.

Before giving the proof, we need to extend the theory of the resultant slightly in such a way that it applies to s polynomials $f_1, \dots, f_s$ rather than just to two. To do so, we introduce new indeterminates $U_2, \dots, U_s$ and regard

$$F = U_2 f_2 + \cdots + U_s f_s$$

as a member of $K[U_2, \dots, U_s, X_1, \dots, X_n]$ whose degree $\deg_1 F$ in X_1 is the maximum of the degrees of $f_2, \dots, f_s$ in X_1. We can then view f_1 as a member of the same polynomial ring $K[U_2, \dots, U_s, X_1, \dots, X_n]$ of degree $\deg_1 f_1$ and form the resultant of f_1 and F in the X_1 variable. This is computed as the determinant of some square matrix of size $\deg_1 f_1 + \deg_1 F$, and we are interested only in the case that $\deg_1 f_1 \geq 1$ and $\deg_1 F \geq 1$. When expanded in monomials $U^\alpha = U_2^{\alpha_2} \cdots U_s^{\alpha_s}$, the determinant is of the form

$$R(f_1, F) = \sum_\alpha h_\alpha(X_2, \dots, X_n) U^\alpha$$

with each h_α in $K[X_2, \dots, X_n]$. The polynomials h_α will be called the **generalized resultants** in the X_1 variable of the ordered pair $(f_1, \{f_2, \dots, f_s\})$.

Proof of Theorem 8.31. Let us abbreviate $\overline{X} = (X_2, \dots, X_n)$ and $\bar{c} = (c_2, \dots, c_n)$; we shall write

$$(X_1, \overline{X}) = (X_1, \dots, X_n) \qquad \text{and} \qquad (X_1, \bar{c}) = (X_1, c_2, \dots, c_n).$$

We seek $c_1 \in K$ with $f_j(c_1, c) = 0$ for all j. The assumption is that $g_i(\bar{c}) \neq 0$ for some i, and we may as well assume that this i is $i = 1$. If $\deg_1 f_1 = 0$, then f_1 is in J, and the conditions that $f_1 = 0$ on $V_K(J)$ and that $g_1(\bar{c}) \neq 0$ contradict one another; hence $\deg_1 f_1 \geq 1$.

As in the paragraph before the proof, put $F = U_2 f_2 + \cdots + U_s f_s$. If $\deg_1 F = 0$, then f_j is independent of X_1 for all $j \geq 2$, and hence f_j is in J for $j \geq 2$. In this case it is enough to find c_1 with $f_1(c_1, \bar{c}) = 0$. Since $g_1(\bar{c}) \neq 0$, $f_1(X_1, \bar{c})$ is a one-variable polynomial of degree $l_1 \geq 1$, and it is 0 for some value c_1. Thus the proof is complete if $\deg_1 F = 0$.

We may therefore assume that $\deg_1 F \geq 1$. Form the resultant in X_1 given by

$$R(f_1, F) = \sum_\alpha h_\alpha(\overline{X}) U^\alpha,$$

where the h_α's are the generalized resultants mentioned above. The main step is to prove that each h_α lies in the first elimination ideal J. Since h_α depends only on $\overline{X}$, it is enough to prove that each h_α is in I. We have arranged that each of f_1

and F has positive degree and has nonzero leading coefficient in X_1, and hence Theorem 8.1 shows that

$$af_1 + bF = R(f_1, F)$$

for some nonzero polynomials a and b in $K[U_2, \dots, U_s, X_1, \overline{X}]$. Let the monomial expansions of a and b in terms of the U^α's be $a = \sum_\alpha a_\alpha U^\alpha$ and $b = \sum_\alpha b_\alpha U^\alpha$. Then we have

$$\sum_\alpha a_\alpha f_1 U^\alpha + \big(\sum_\beta b_\beta U^\beta\big)\big(\sum_{i=2}^{s} f_i U_i\big) = \sum_\alpha h_\alpha U^\alpha. \qquad (*)$$

Let e_i be the multi-index that is 1 in the i^{th} place and 0 elsewhere. This has the property that $U^{e_i} = U_i$ for $2 \le i \le s$. We can rewrite $(*)$ as

$$\sum_\alpha h_\alpha U^\alpha = \sum_\alpha a_\alpha f_1 U^\alpha + \sum_\alpha \Big(\sum_{\substack{(\beta,i) \text{ with} \\ 2 \le i \le s, \\ \beta + e_i = \alpha}} b_\beta f_i \Big) U^\alpha.$$

Equating the coefficients of U^α on both sides gives

$$h_\alpha = a_\alpha f_1 + \sum_{\substack{(\beta,i) \text{ with} \\ 2 \le i \le s, \\ \beta + e_i = \alpha}} b_\beta f_i$$

and exhibits h_α as in I. Therefore h_α is in the elimination ideal J.

Since $\bar{c}$ lies in $V_K(J)$, $h_\alpha(\bar{c}) = 0$ for all α. Consequently

$$R(f_1, F)(U_2, \dots, U_s, \bar{c}) = 0.$$

Theorem 8.1 shows that $f_1(X_1, \bar{c})$ and $F(U_2, \dots, U_s, X_1, \bar{c})$ have a common factor of positive degree in X_1 provided either or both of two specific coefficients are nonzero. These are the coefficients of $X_1^{\deg_1 f_1}$ in $f_1(X_1, \bar{c})$ and of $X_1^{\deg_1 F}$ in $F(U_2, \dots, U_s, X_1, \bar{c})$. The coefficient of $X_1^{\deg_1 f_1}$ in $f_1(X_1, \overline{X})$ is $g_1(\overline{X})$; thus the coefficient of $X_1^{\deg_1 f_1}$ in $f_1(X_1, \bar{c})$ is $g_1(\bar{c})$ and is nonzero by assumption. Therefore Theorem 8.1 is applicable.

The common factor of $f_1(X_1, \bar{c})$ and $F(U_2, \dots, U_s, X_1, \bar{c})$ may be taken to be prime, and then it has to be a nonzero scalar multiple of $X_1 - c_1$ for some $c_1 \in K$, since that is the only kind of prime factor that divides $f_1(X_1, \bar{c})$, K being algebraically closed. Thus the element c_1 of K satisfies

$$f_1(c_1, \bar{c}) = 0 \qquad \text{and} \qquad F(U_2, \dots, U_s, c_1, \bar{c}) = 0. \qquad (**)$$

Writing out F, we have

$$0 = F(U_2, \dots, U_s, c_1, \bar{c}) = U_2 f_2(c_1, \bar{c}) + \cdots + U_s f_s(c_1, \bar{c}).$$

This is an identity in $K[U_2, \dots, U_s]$, and each coefficient must be 0 on the right side. Thus $0 = f_2(c_1, \bar{c}) = \cdots = f_s(c_1, \bar{c})$. Since $(**)$ shows that $f_1(c_1, \bar{c}) = 0$, this proves the theorem. □

11. Problems

1. How many points are in $\mathbb{P}^n_K$ if K is a finite field with q elements?
2. Resolve Cramer's paradox as formulated in Section 1.
3. **(Euler's Theorem)** Prove that if $F(X_1, \dots, X_n)$ is any homogeneous polynomial of degree d, then $\sum_{j=1}^n X_j \frac{\partial F}{\partial X_j} = dF$.
4. Let A and B be unique factorization domains, and let $\iota : A \to B$ be a one-one homomorphism of commutative rings with identity. For each $h(X)$ in $A[X]$, let $h^\iota(X)$ be the member of $B[X]$ obtained by applying the substitution homomorphism that acts by ι on the coefficients and fixes X. Using resultants, prove that if $f(X)$ and $g(X)$ are two members of $A[X]$ such that $f^\iota(X)$ and $g^\iota(X)$ have a common factor in $B[X]$ that is not in B, then f and g have a common factor in $A[X]$ that is not in A.
5. Theorem 8.1 assumes that at least one of the coefficients f_m and g_n is nonzero. Sometimes this theorem is phrased with the stronger hypothesis that f_m and g_n are both nonzero. By comparing the resultants that are involved, show that all parts of the theorem with at least one of f_m and g_n nonzero are consequences of the theorem with both f_m and g_n nonzero.
6. Let K be an algebraically closed field, let f and g be members of $K[X_1, \dots, X_n]$ with f irreducible, and suppose that $g(a_1, \dots, a_n) = 0$ whenever $f(a_1, \dots, a_n) = 0$. Give two proofs, one using the Nullstellensatz and one using resultants, that f divides g.
7. Factor the member $Y^3 - 2XY^2 + 2X^2Y - 4X^3$ of $\mathbb{C}[X, Y]_3$ into first-degree factors.
8. Find the intersections in $\mathbb{P}^2_{\mathbb{C}}$ of the zero loci of the projective plane curves $F(X, Y, W) = X(Y^2 - XW)^2 - Y^5$ and $G(X, Y, W) = Y^4 + Y^3W - X^2W^2$.
9. Let A be a unique factorization domain, let $B = A[Y_1, \dots, Y_m, Z_1, \dots, Z_n]$, let F and G be the polynomials in $B[X]$ given by

$$F(X) = \prod_{i=1}^m (X - Y_i) \qquad \text{and} \qquad G(X) = \prod_{j=1}^n (X - Z_j),$$

and let $R(Y_1, \dots, Y_m, Z_1, \dots, Z_n)$ be the resultant $R(F, G)$ with respect to X.

(a) Show that $R(Y_1, \dots, Y_m, Z_1, \dots, Z_n)$ equals 0 if Y_i is set equal to Z_j.

(b) Deduce from (a) that $Y_i - Z_j$ divides $R(Y_1, \dots, Y_m, Z_1, \dots, Z_n)$.

(c) Deduce from (b) that $R(Y_1, \dots, Y_m, Z_1, \dots, Z_n) = c \prod_{i,j} (Y_i - Z_j)$ for some $c \neq 0$ in A depending on m and n.

10. Let $f(X)$ be in $K[X]$, K being a field, and let $f'(X)$ be the derivative of $f(X)$. Using the result of the previous problem and the computation at the beginning of Section V.4, prove that $R(f, f')$ is a nonzero multiple of the discriminant of f, the multiple depending only on $\deg f$.

11. Let F and G be the homogeneous polynomials given by $F(X, Y, W) = (X^2 + Y^2)^2 + 3X^2YW - Y^3W$ and $G(X, Y, W) = (X^2 + Y^2)^3 - 4X^2Y^2W^2$. Calculate $I(P, F \cap G)$ for $P = [0, 0, 1]$.

12. Let G be a nonconstant homogeneous polynomial in $K[X, Y, W]_d$ vanishing at a point P of $\mathbb{P}^2_K$, let $m = m_P(G)$ be the order of vanishing of G at P, and let L be a projective line through P. Show from the definitions that L is a tangent line to G at P in the sense of Section 5 if and only if $i(P, L \cap G) \geq m + 1$ in the sense of Section 4.

13. Deduce relative to an arbitrary monomial ordering the (nonconstructive) existence of a Gröbner basis for a nonzero ideal I in $K[X_1, \dots, X_n]$ from the form of a set of generators of the ideal $\mathrm{LT}(I)$.

14. For $1 \leq i \leq n$, let $w^{(i)}$ be the weight vector $w^{(i)} = (w_1^{(i)}, \dots, w_n^{(i)})$ in $\mathbb{R}^n$, and suppose that these vectors are linearly independent. Show that the $w^{(i)}$ define a monomial ordering as in Example 5 of Section 7 if and only if for each j, the first i with $w_j^{(i)} \neq 0$ has $w_j^{(i)} > 0$.

15. This problem shows for two variables that every monomial ordering arises from a system of two independent weight vectors satisfying the condition in the previous problem. Let a monomial ordering be imposed on $K[X, Y]$.

(a) If $X > Y^q$ for all $q > 0$, show that the ordering is lexicographic and is determined by the system of two weight vectors $\{(1, 0), (0, 1)\}$.

(b) If $X < Y^q$ for some $q > 0$, show that there exists a unique real number $r \geq 0$ such that for all ordered pairs of integers $u \geq 0$ and $v \geq 0$, $X^u > Y^v$ if $ru > v$ and $X^u < Y^v$ if $ru < v$.

(c) If $X < Y^q$ for some $q > 0$ and if r is defined as in (b), prove that the monomial ordering is determined by the system of two weight vectors $\{(r, 1), (s, t)\}$ for a suitable (s, t).

16. In $K[X, Y]$, define $f(X, Y) = X^2Y + XY^2 + Y^2$, $f_1(X, Y) = XY - 1$, and $f_2(X, Y) = Y^2 - 1$. Show that

$$f(X, Y) = (X + Y)f_1 + 1f_2 + r_1 = Xf_1 + (X + 1)f_2 + r_2$$

with $r_1(X, Y) = X + Y + 1$ and $r_2 = 2X + 1$ gives two decompositions in the lexicographic ordering of f relative to $\{f_1, f_2\}$ satisfying the conditions of the generalized division algorithm of Proposition 8.20. Conclude that the remainder term need not be unique, nor need the coefficients of f_1 and f_2.

17. Observe for any scalar a that the ideal $I = (X^2 + cXY, XY)$ in $K[X, Y]$ is independent of c.
    (a) Verify that $\{X^2 + cXY, XY\}$ is a minimal Gröbner basis of I relative to the lexicographic ordering for any choice of c.
    (b) Show that $\{X^2, XY\}$ is the reduced Gröbner basis for I.

Problems 18–20 characterize ideals in $K[X_1, \dots, X_n]$ whose locus of common zeros is a finite set under the assumption that K is an algebraically closed field. Thus let K be an algebraically closed field, and let I be a nonzero ideal in $K[X_1, \dots, X_n]$.

18. Under the assumption for each j with $1 \leq j \leq n$ that I contains a nonconstant polynomial $P_j(X_j)$, prove that $V_K(I)$ is a finite set.

19. Conversely under the assumption that $V_K(I))$ is a finite set, use the Nullstellensatz to produce for each j, a nonconstant polynomial $P_j(X_j)$ lying in I.

20. Impose the usual lexicographic ordering on monomials. Prove that $\mathrm{LT}(I)$ contains some $X_j^{l_j}$ for each j with $1 \leq j \leq n$ if and only if $V_K(I)$ is a finite set. (Educational note: The advantage of this characterization over the one in Problems 18–19 is that checking this one is easy by inspection once a Gröbner basis of I has been computed.)

Problems 21–23 relate solutions of simultaneous systems of polynomial equations to the theory of the Brauer group in Chapter III. A field L is said to satisfy **condition (C1)** if every homogeneous polynomial of degree d in n variables with $d < n$ has a nontrivial zero. The significance of this condition was shown in Problem 20 at the end of Chapter III: the Brauer group $\mathcal{B}(L)$ of such a field is necessarily 0. The present set of problems establishes that a simple transcendental extension of an algebraically closed field satisfies condition (C1). No knowledge of Chapter III is needed for these problems, but Problem 23 will take for granted a certain theorem to be proved in Chapter X.

21. Let K be an algebraically closed field, and let $L = K(X)$ be a simple transcendental extension. It is to be shown that any member $F(T_1, \dots, T_n)$ of $L[T_1, \dots, T_n]_d$ of the form $F(T_1, \dots, T_n) = \sum_{i_1,\dots,i_n} a_{i_1\cdots i_n} T_1^{i_1} \cdots T_n^{i_n}$ has a nontrivial zero if $d < n$ and each $a_{i_1,\dots,i_n}$ lies in the field $L = K(X)$.
    (a) Why is it enough to consider such polynomials with each $a_{i_1,\dots,i_n}$ in the polynomial ring $K[X]$?

(b) With the simplification from (a) in place, let δ be the maximum degree in X of the coefficients $a_{i_1\cdots i_n}$. Let N be a positive integer to be specified. By looking for a solution of the form $T_i = \sum_{j=0}^{N} b_{ij} X^j$ with each b_{ij} in K, show that substitution of this formula into the formula $F(T_1, \dots, T_n) = 0$ leads to a system of homogeneous polynomial equations over K in the unknowns b_{ij}, one of each degree from 0 to $\delta + Nd$.

22. (a) In the setting of the previous problem, show that the number of unknowns is $(N+1)n$ and that the number of equations is at most $Nd + \delta + 1$.

(b) Show for N sufficiently large that the number of equations is less than the number of unknowns.

23. The following theorem will be discussed in Chapter X: if K is algebraically closed and if $m \leq n$, then the locus of common zeros in $\mathbb{P}^n_K$ of m nonconstant homogeneous polynomials in $K[X_1, \dots, X_{n+1}]$ is nonempty. Assuming this theorem, deduce from the previous two problems the conclusion that the field $L = K(X)$ satisfies condition (C1) if K is algebraically closed.