Linear Systems of Plane Curves

Rick Miranda

Introduction

Interpolation with polynomials is a subject that has occupied mathematicians' minds for millenia. The general problem can be informally phrased as: Given a set of points $\{(x_i, y_i)\}$ in the plane, find a polynomial $f(x, y)$ with these points as roots. (The one-variable version of this problem is easy and crops up in the high school and undergraduate curriculum on occasion.) Sometimes one asks for a polynomial of minimal degree that works. The condition to pass through a point is a linear one, and so a more sophisticated version of the interpolation problem asks for the vector space of *all* interpolating polynomials whose degree is bounded by a particular integer.

In this article I want to draw the reader's attention to the *dimension* of such a space. It turns out that if the points are in very special positions, the interpolation conditions may be dependent, which complicates the problem, of course. We then make the blanket assumption that the points are in *general* position, which insures that the interpolation conditions are independent; but then, as one realizes immediately, the dimension question is trivial: the codimension of the space of interpolating polynomials is equal to the number of points being imposed.

We therefore recomplicate the problem by requiring a more subtle form of interpolation: for each point p we fix an integer m (depending on p) and require that the polynomial not only vanish at p, but vanish "to order m": it and all of its derivatives up through order $m - 1$ also vanish at p. We often say that the polynomial has *multiplicity* m at p if it vanishes to order m.

This more general problem, even for points in general position, turns out to be surprisingly complicated. The number of conditions imposed by asking a polynomial $f(x, y)$ to vanish to order m at p is $m(m+1)/2$: this is just the number of terms in the Taylor expansion of f at p up through order $m - 1$, and all of these coefficients must vanish. So the naive conjecture would be that the codimension of the space of interpolating polynomials vanishing to order m_i at p_i is equal to $\sum_i m_i(m_i + 1)/2$ (unless, of course, the interpolating space is empty); this gives the "expected" dimension, assuming that all the linear conditions being imposed are independent.

This naive conjecture is false, as some easy examples show. The simplest is to consider the space of conics having multiplicity two at each of two points p and q. The vector space dimension of conics is six, and each multiplicity-two point gives three conditions, so one expects there to be *no* nonzero conics double at the two points. However, if $e(x, y)$ is the linear polynomial defining the line through p and q, then $e(x, y)^2$ is a nonzero conic double at p and q.

The real conjecture, due to Harbourne and Hirschowitz in the mid-1980s, states that if the space of interpolating polynomials does not have the expected dimension, then there is a polynomial $e(x, y)$ such that every interpolating polynomial is divisible by $e(x, y)^2$. Moreover, the polynomial $e(x, y)$ has a special form, which will be exposed later.

Although the conjecture is rather recent, the general problem goes back to the last century, more or less to the origins of algebraic geometry. Bézout

---

Rick Miranda is professor of mathematics at Colorado State University. His e-mail address is `miranda@math.colostate.edu`.

in the 1700s; Plücker, Bertini, and M. Noether in the 1800s; and Terracini, Castelnuovo, Segre, and Severi in this century all made contributions, among many others. Coolidge's treatise [Co] makes for interesting reading even today.

My purpose in writing this article is to pique the interest of *Notices* readers in this curious situation and to explain some of the methods leading to the conjecture. Along the way I shall discuss some of the variations and applications of the conjecture (in particular, the higher-dimensional analogue of the problem and its relationship to Waring's problem for forms) and give an impression of some recent results.

Basic Definitions

The Projective Space of Plane Curves

We regard the underlying field as the complex numbers. Informally, an algebraic curve in the plane is given by the vanishing of a nonzero polynomial $f(x, y) = 0$. This is a bit dangerous if we want to focus on the polynomials themselves, since the set of zeroes of the polynomial does not determine the polynomial: for example, f and f^2 have the same zeroes. The only ambiguity not affecting later computations is multiplication of the polynomial by a nonzero constant. We therefore define an algebraic plane curve (of degree d) to be an equivalence class of nonzero polynomials of degree at most d, two being identified if they are scalar multiples of one another. Such polynomials, including the zero polynomial, form a vector space of dimension $(d^2 + 3d + 2)/2$, which is the number of monomials in x and y of degree at most d.

With this definition, we see that the space of plane curves is naturally a *projective space*, that is, the space of 1-dimensional subspaces of the vector space of polynomials of degree at most d. The dimension of a projective space is always one less than that of the corresponding vector space. A single point has dimension 0 and the empty set, projectively, has dimension -1. The projective space of 1-dimensional subspaces of $\mathbb{C}^{r+1}$ is denoted by $\mathbb{P}^r$.

Therefore the projective space $\mathcal{L}_d$ of plane curves of degree d has dimension $d(d + 3)/2$.

A *component* of a curve C, with defining polynomial f, is the curve associated to any nonconstant factor of f; a curve is *irreducible* if its defining polynomial cannot be factored. Geometrically every curve is the union of its irreducible components.

Plane Curves Interpolating Points with Multiplicity

Fix a point p in the plane, and denote by $\mathcal{L}_d(-mp)$ the set of plane curves of degree d which have multiplicity at least m at p. Having multiplicity at least one at p is equivalent to the curve passing through p.

As noted above, for a curve C defined by $f(x, y) = 0$ to have multiplicity at least m at p is exactly $m(m + 1)/2$ independent linear conditions on the coefficients of f, as long as the degree of f is at least $m - 1$. Therefore the set $\mathcal{L}_d(-mp) \subset \mathcal{L}_d$ is either empty (if $d \le m - 1$) or has codimension $m(m + 1)/2$.

Suppose now that $p_1, \dots , p_n$ are distinct points in the plane and that $m_1, \dots , m_n$ are nonnegative integers. Define

$$\mathcal{L}_d\left(-\sum_{i=1}^{n} m_i p_i\right) = \{C \in \mathcal{L}_d(-m_i p_i) \text{ for every } i\} = \bigcap_{i=1}^{n} \mathcal{L}_d(-m_i p_i)$$

to be the space of plane curves of degree d having multiplicity at least m_i at p_i for each i. This is a linear subspace of $\mathcal{L}_d$, called a *linear system* of plane curves. The problem we have set ourselves to investigate is, What is the dimension of the linear system $\mathcal{L}_d(-\sum_{i=1}^{n} m_i p_i)$?

The Virtual and Expected Dimension

As noted above, one knows that for each point p the $m(m + 1)/2$ conditions imposed by asking the curve to have multiplicity at least m are independent. However, if there is more than one point, it is not at all clear whether the totality of all these conditions at every point is independent.

If so, we have an easy formula for the dimension: if there are n points, define the *virtual dimension* of $\mathcal{L} = \mathcal{L}_d(-\sum_{i=1}^{n} m_i p_i)$ by

$$\nu = \nu_d\left(-\sum_{i=1}^{n} m_i p_i\right) = \frac{d(d+3)}{2} - \sum_{i=1}^{n} m_i(m_i + 1)/2. \tag{1}$$

Note that if this number is negative, then we really expect the space $\mathcal{L}$ to be empty; hence we define the *expected dimension* of $\mathcal{L}$ to be

$$e = e_d\left(-\sum_{i=1}^{n} m_i p_i\right) = \max\{-1, \nu\}. \tag{2}$$

We note that

$$\dim \mathcal{L}_d\left(-\sum_{i=1}^{n} m_i p_i\right) \ge e_d\left(-\sum_{i=1}^{n} m_i p_i\right) \ge \nu_d\left(-\sum_{i=1}^{n} m_i p_i\right),$$

since the failure of the conditions to be independent can only raise the dimension of the interpolating space. Notice that the second and third quantities are equal if either is at least -1.

Whether this space of curves has the expected dimension depends on the positions of the points, even if all of the multiplicities are one.

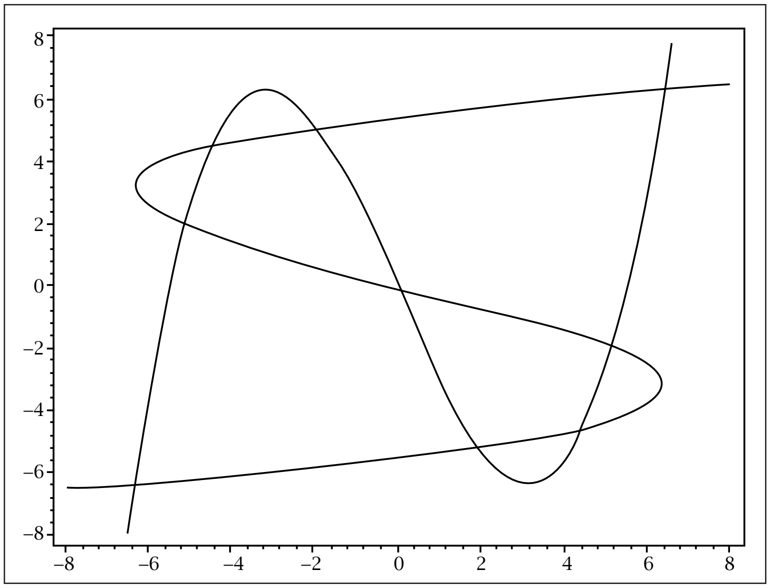

Figure 1. The nine intersections of two cubic curves, namely, $x = .1(y^3 - 30y)$ and $y = .1(x^3 - 30x)$. Only points with real coordinates are plotted, and all nine intersections are of this kind. For these cubics, there are no intersections at the points at infinity in the projective plane, and the intersection numbers of the two curves are 1 at each point of intersection.

Example 1. Suppose $d + 2 \le n \le d(d+3)/2$ and all the points p_i, $1 \le i \le n$, lie on a line ℓ. We can change coordinates and assume that the line is the x-axis, given by $y = 0$, and hence the points p_i have the form $p_i = (x_i, 0)$. Suppose that C is a plane curve of degree d containing $d+1$ of the points, say $p_1, \dots, p_{d+1}$. If $f(x, y) = 0$ defines C, then this means that $f(x_i, 0) = 0$ for $1 \le i \le d+1$, and therefore the polynomial $f(x, 0)$ has $d+1$ roots. Since it is a polynomial of degree at most d in x, it must be identically zero. The only way $f(x, 0)$ can be identically zero is if every term of $f(x, y)$ contains y, which is equivalent to having y divide f. If this happens, then the curve C contains the line ℓ as a component.

What we have shown is that any curve $C \in \mathcal{L}_d(-\sum_{i=1}^{d+1} p_i)$ must, in this case, contain the line ℓ through all of the n points. Therefore, any curve of degree d containing the first $d+1$ points automatically contains the remaining $n - d - 1$ points. Hence $\mathcal{L}_d(-\sum_{i=1}^{n} p_i) = \mathcal{L}_d(-\sum_{i=1}^{d+1} p_i)$, so that

$$\begin{aligned}\dim \mathcal{L}_d\left(-\sum_{i=1}^{n} p_i\right) &\\ &= \dim \mathcal{L}_d\left(-\sum_{i=1}^{d+1} p_i\right) \ge v_d\left(-\sum_{i=1}^{d+1} p_i\right) \\ &= \frac{d(d+3)}{2} - (d+1) > \frac{d(d+3)}{2} - n \\ &= e_d\left(-\sum_{i=1}^{n} p_i\right),\end{aligned}$$

the final equality holding because $n \le d(d+3)/2$. Hence we see in this case that the actual dimension is strictly greater than the expected dimension.

Example 2. Let C and D be two plane cubic curves which intersect nine times (this is forced by Bézout's theorem, which we shall discuss below). Suppose in fact they intersect transversally (crossing each other with distinct tangents) at nine distinct points $p_1, \dots, p_9$ (which happens generically, in fact). Consider the linear system $\mathcal{L} = \mathcal{L}_3(-\sum_{i=1}^{9} p_i)$ of cubics through the nine points. Since we have by construction two such cubics, this linear space contains two elements and so must have dimension at least one. But $v = 9 - 9 = 0$, so that the expected dimension e is also zero. Hence, we again have the situation where the actual dimension is greater than the expected dimension. See Figure 1.

Choosing Points Generically

The previous examples have dimension greater than the expected dimension because the positions of the points are special in some way. To understand this phenomenon, it is useful to consider the linear system $\mathcal{L}_d(-\sum_i m_i p_i)$, as a vector space of polynomials, to be the kernel of a linear map ϕ. This map ϕ simply takes a polynomial f to a t-tuple (where $t = \sum_i m_i(m_i+1)/2$) by evaluating f and all relevant derivatives at the points p_i. The map ϕ has a matrix when we use the basis of monomials $\{x^r y^s\}$ for the vector space of all polynomials of degree at most d. In this basis, if $p_i = (x_i, y_i)$, then the matrix for ϕ has entries equal to the derivatives of the monomials $x_i^a y_i^b$.

Let $\mathcal{R}_k$ be the set of all n-tuples of points $(p_1, \dots, p_n)$ where the above matrix has rank at most k. We clearly have $\mathcal{R}_{k-1} \subseteq \mathcal{R}_k$ for every k; moreover, for large k we see that $\mathcal{R}_k = (\mathbb{P}^2)^n$, since as soon as k is larger than the size of the matrix, there is no condition on the points p_i. Therefore, there is a maximum number r such that $\mathcal{R}_r \ne (\mathbb{P}^2)^n$, but $\mathcal{R}_{r+1} = (\mathbb{P}^2)^n$.

Suppose that the points $\{p_i\}$ are chosen so that $(p_1, \dots, p_n\}$ does *not* lie in $\mathcal{R}_r$. Then the matrix for ϕ has the maximum possible rank (namely, $r+1$) so that the dimension of the kernel has minimum possible dimension. The locus $\mathcal{R}_r$ is a closed subset of $(\mathbb{P}^2)^n$, defined by the vanishing of all $(r+1) \times (r+1)$ minors of the matrix for ϕ, and has dimension strictly smaller than $2n$, which is the dimension of $(\mathbb{P}^2)^n$. Therefore, we see that if we choose the n-tuple of points off this closed subset of lower dimension, the space $\mathcal{L}_d(-\sum_i p_i)$ achieves its minimum possible dimension, which we call the *generic dimension* for curves of degree d having the required multiplicities at the n points. For this problem, n-tuples of points off this closed subset will be said to be *in general position*.

The Multiplicity One Theorem

We can now address the situation that arose first in the introduction: the simple interpolation of polynomials, without higher multiplicities. In this case there are no surprises; all such systems have the expected dimension.

Multiplicity One Theorem. If the points $\{p_i\}$ are in general position, then the dimension of $\mathcal{L}_d(-\sum_i p_i)$ is equal to the expected dimension.

Proof: We prove this by induction on the number of points n. For $n = 0$ it is clear, and for $n = 1$ we are claiming that we can choose the point p_1 so that it does not lie on every curve of degree d, which is obvious. For the general induction step, since each additional point contributes exactly one linear condition, all that is required is to show that this condition is not dependent on the previous conditions. This is equivalent to having each additional point not lie on every curve passing through the previous points. Of course, if the points are in general position, it will not: simply take any curve passing through the previous points and take as the additional point any point not on that curve. □

First Examples of Special Systems

We say that a linear system $\mathcal{L}_d(-\sum_i m_i p_i)$ is *special* if it does not have the expected dimension; otherwise it is *nonspecial*. The Multiplicity One Theorem says that if all m_i are equal to one, then the system is nonspecial. A naive conjecture would be the analogue of the Multiplicity One Theorem: For generic choices of the points, the dimension of the system is equal to the expected dimension. As noted in the introduction, this is false, as the example of conics double at two points shows: $\dim \mathcal{L}_2(-2p - 2q) = 0$, but the expected dimension is -1.

Let us offer another example of a case where no matter how general the choice of the points is, the linear system is special.

Example 3. Choose five points $p_1, \ldots, p_5$ in the plane generically. Then $\dim \mathcal{L}_2(-\sum_i p_i) = 0$ by the Multiplicity One Theorem, since the virtual dimension is $5 - 5 = 0$: there is a unique conic through the five points. Let e be the quadratic equation defining this conic, and let $f = e^2$. Then f is a quartic, and as above the multiplicity of f is 2 at all points of the conic, in particular at each of the original five points. Hence $\mathcal{L}_4(-\sum_i 2p_i)$ is not empty. However, its virtual dimension is $v = 14 - 5 \cdot 3 = -1$, and so the system is expected to be empty.

(–1)-Curves

Intersection Multiplicities

To begin to understand the phenomenon of special systems, it is necessary to introduce the notion of intersection multiplicities between two curves. Briefly speaking, if two curves C_1 and C_2 meet at a point p, we want to carefully count *how many times* they meet there. This is nothing more than a generalization of the multiplicity of a root of a single equation in one variable; here we have two equations in two variables. I do not want to enter into the technicalities in this article. One can consult [F] for an elementary treatment; suffice it to say that if the curves C_1 and C_2 do not have a common component passing through a point p, then a nonnegative integer $I_p(C_1, C_2)$ is defined that measures the "intersection multiplicity" we need. It satisfies the following properties:

- $I_p(-,-)$ is bilinear over the integers,
- $I_p(C_1, C_2) \geq 1 \iff p \in C_1$ and $p \in C_2$,
- $I_p(C_1, C_2) \geq \text{mult}_p(C_1) \cdot \text{mult}_p(C_2)$, and
- $\sum_p I_p(C_1, C_2) = \deg(C_1) \cdot \deg(C_2)$.

The first three of these properties are not hard to understand. The first is quite natural: it says that if C meets D a total of r times at p and meets E a total of s times at p, then it meets the union $D + E$ a total of $r + s$ times at p. (Here the "sum" of two curves is given by the product of their defining functions.) The second simply indicates that $I_p = 0$ when p is not a root of both equations. The third generalizes the second: a curve C having multiplicity m at p is, generically, m branches, or separate smooth arcs, through p. If C_i has multiplicity m_i at p, then $I_p(C_1, C_2)$ gets a contribution of at least one from each branch of C_1 meeting each branch of C_2.

The final property is more subtle and is *Bézout's Theorem*: the number of intersections is the product of the degrees, "counted properly". Here counting properly means using the intersection multiplicity, and the sum is taken over all points in the *projective plane*, including, of course, whatever points at infinity are involved in the intersection. Bézout's Theorem holds only when the curves do not share a common component. It is the generalization to two variables of the fact that a polynomial in one variable of degree d has exactly d roots, counting roots according to their order.

Extra Intersections

Fix once and for all points $p_1, \ldots, p_n$ in the plane. Consider a curve $D \in \mathcal{L}_d(-\sum m_i p_i)$ and another curve $E \in \mathcal{L}_e(-\sum k_i p_i)$: where do D and E meet? At one of the chosen points p_i, D has multiplicity at least m_i and E has multiplicity at least k_i, so that the third property of intersection numbers says that $I_{p_i}(D, E) \geq m_i k_i$. Therefore we have at least $\sum_i m_i k_i$ intersections all accounted for at the chosen points. The total number of intersections is, by Bézout's Theorem, the product de of the degrees. Therefore it is natural to consider the number of "extra intersections" $de - \sum_i m_i n_i$.

This quantity has several advantages, the primary one being that it is *bilinear* (over the integers) in the given data (of degrees and multiplicities); in-

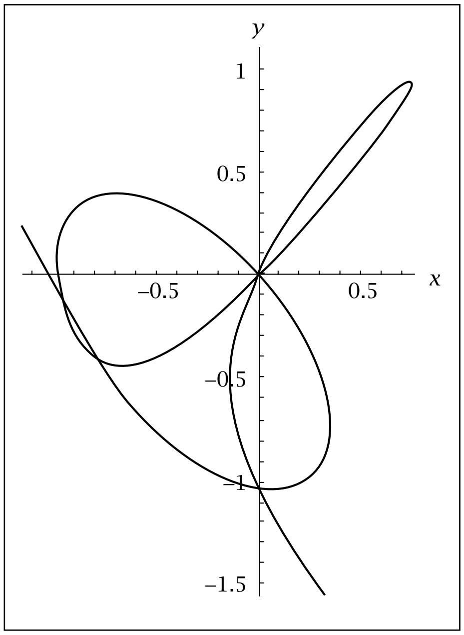

Figure 2. A quintic with one triple point and three other double points. The equation is:
$18x^5+5x^4y-9x^3y^2-17x^2y^3+2xy^4$
$+9y^5+36x^4-4x^3y-29x^2y^2-16xy^3$
$+18y^4+18x^3-9x^2y-18xy^2+9y^3=0$.
The curve has a triple point at (0,0) and double points at (–1,0), (0,–1), and (–18/23, –9/23).

deed, it depends only on these data. We are inexorably led to giving these data lives of their own: let us define a *curve class* to be an $(n+1)$-tuple of integers $(d; m_1, m_2, \ldots, m_n)$; these form a group under addition. If C is an actual curve in the plane, we say that C belongs to the curve class $(d; m_1, m_2, \ldots, m_n)$ if $\deg(C) = d$ and the multiplicity of C at p_i is at least m_i for each i.

Curve classes have associated linear systems (the space of curves $\mathcal{L}_d(-\sum_i m_i p_i)$), and derived from this they have virtual, expected, and actual dimensions. The first two of these are defined by (1) and (2) as earlier. The third is the generic dimension of $\mathcal{L}_d(-\sum_i m_i p_i)$ determined when the points are in general enough position.

Given two curve classes, we may define their *extra-intersection number*, motivated by the discussion above: if $\mathbf{d} = (d; m_1, \ldots, m_n)$ and $\mathbf{e} = (e; n_1, \ldots, n_n)$ are two curve classes, define $\langle \mathbf{d}, \mathbf{e} \rangle = de - \sum_i m_i n_i$. This is a bilinear function of curve classes.

One immediate result is that if C and D are curves belonging to curve classes $\mathbf{c}$ and $\mathbf{d}$ respectively, and $\langle \mathbf{c}, \mathbf{d} \rangle < 0$, then C and D must share a common component, since Bézout's Theorem is failing to hold. If, for example, C is an irreducible curve, then in this case C must be a component of D. We will often abuse the notation for the extra-intersection number slightly by applying it to actual curves instead of the curve classes; in this way we would write $\langle C, D \rangle$ for $\langle \mathbf{c}, \mathbf{d} \rangle$. We occasionally abuse the notation further and apply it to the linear systems $\mathcal{L}_d(-\sum_i m_i p_i)$ also.

The Genus Formula

Define the "canonical curve class" $\mathbf{k} = (-3; -1, -1, \ldots, -1)$; this has all negative data, but the opposite curve class $-\mathbf{k} = (3, 1, 1, \ldots, 1)$ is more easily understood: its linear system consists of cubic curves passing through all the points. There may not be any such curves, but the algebra of extra intersections does not really care.

With this class in hand, we have a formula for the virtual dimension of a curve class $\mathbf{c} = (d; m_1, \ldots, m_n)$, expressed entirely in terms of extra-intersection numbers:

$$\begin{aligned} \nu(\mathbf{c}) &= \nu_d\Big(-\sum_i m_i p_i\Big) \\ &= \frac{d(d+3)}{2} - \sum_i \frac{m_i(m_i+1)}{2} \\ &= \langle \mathbf{c}, \mathbf{c} - \mathbf{k} \rangle / 2. \end{aligned} \tag{3}$$

Here the difference $\mathbf{c} - \mathbf{k}$ is the class with the difference of the data: $\mathbf{c} - \mathbf{k} = (d+3; m_1+1, \ldots, m_n+1)$. This formula happens to be a special case of the celebrated Riemann-Roch Theorem, but this fact will not concern us.

So far this extra-intersection number has not been useful in telling us anything we did not know already; let us remedy this. Recall that a smooth complex curve is a *Riemann surface*, i.e., a closed one-dimensional complex manifold; as such it is a closed orientable real two-manifold, and topologically it may be described by its *genus*.

A sphere has genus zero: examples are lines and conics in the plane. A torus has genus one: a smooth plane cubic curve is the prototype. If a curve C belonging to a curve class $\mathbf{c}$ is smooth and irreducible after resolving the singularities that are forced at the chosen points p_i with $m_i \geq 2$, then it will also have a genus g. The Genus Formula in the case that all of the singularities are ordinary (that is, each point of multiplicity m consists of m branches with distinct tangents) is

$$g(C) = 1 + \langle C, C + \mathbf{k} \rangle / 2, \tag{4}$$

which expresses the genus of C in terms of an extra-intersection number. The formula actually decomposes into

$$g(C) = (d-1)(d-2)/2 - \sum m_i(m_i - 1)/2,$$

which some readers may recognize as a generalization of the famous Plücker formula, one of the cornerstones of surface theory.

For an irreducible curve C, the genus must be nonnegative. If a class is encountered with negative genus (which is numerically possible), its linear system cannot contain irreducible curves.

(–1)-Curves

The *self-intersection* of a class $\mathbf{c}$ is the integer obtained by extra-intersecting $\mathbf{c}$ with itself: $\langle \mathbf{c}^2 \rangle = \langle \mathbf{c}, \mathbf{c} \rangle$. If $\mathbf{c} = (d; m_1, \ldots, m_n)$, then $\langle \mathbf{c}^2 \rangle = d^2 - \sum m_i^2$. It can well happen that a curve class $\mathbf{c}$ has negative self-intersection, even if it is a curve class of a real curve! This may initially be disturbing, but if one thinks for a moment, there is no contradiction to Bézout's Theorem. A simple example is a line through two chosen points, i.e., the curve class $(1; 1, 1)$. Another is a conic through five points; both these examples have $\langle C^2 \rangle = -1$.

These two examples are also extremal in the sense that they are both irreducible curves with zero genus, the minimum possible. Note from (4) that when $\langle C^2\rangle = -1$ and $\langle C, \mathbf{k}\rangle = -1$, then we have a genus zero curve. The condition that $g(C) = 0$ is equivalent to $\langle C, \mathbf{k}\rangle = -1$ if $\langle C^2\rangle = -1$. Such curves, which are simple spheres after resolving the singularities, are called (-1)-*curves*; every interesting extra-intersection number is -1.

Some other interesting (-1)-curves are cubics with one double point, passing through six other points. In general curves of degree e with one point of multiplicity $e-1$ and $2e$ points of multiplicity one are (-1)-curves. These are not the only types: sextics with one triple point and seven double points start another whole family. In general there are infinitely many sets of numerical data giving (-1)-curves, although only finitely many with eight or fewer chosen points. There are already infinitely many with nine points alone.

The Main Conjecture

Generating Special Systems with (–1)-Curves

Recalling two of our first examples of special systems, namely, conics double at two points and quartics double at five points, we see that each of these involve (-1)-curves in an essential way: these systems are expected to be empty, but in fact each contains as its unique element the double of a (-1)-curve. (The *double* of a curve is the curve defined by the square of the defining polynomial.) This turns out to be a critical observation and leads to the systematic generation of other special systems.

Indeed, suppose one wants to explore the generality of this type of example. We seek a curve E that exists whose double $2E$ is not expected to exist. We obtain the following system of inequalities for the relevant extra-intersection numbers:

$$\langle E^2\rangle - \langle E, \mathbf{k}\rangle \geq 0 \quad \text{(E is expected to exist, by (3))}$$

$$\langle E^2\rangle + \langle E, \mathbf{k}\rangle \geq -2 \quad \text{(E has a nonnegative genus, by the Genus Formula)}$$

$$4\langle E^2\rangle - 2\langle E, \mathbf{k}\rangle \leq -1 \quad \text{($2E$ is not expected to exist, by (3)).}$$

The only solution to these inequalities is $\langle E^2\rangle = \langle E, \mathbf{k}\rangle = -1$, forcing E to be a (-1)-curve!

This little back-of-the-envelope calculation is not definitive but at least serves the purpose of focusing our attention on (-1)-curves and the way they prevent linear systems from having the expected dimension. In the above example it is the linear system of the double class $2E$ that visibly exists but is not expected to. It is not only the numerology of the (-1)-curve, but also the fact that it occurs *doubly* in the unexpected system that is important.

To see why this is true, suppose that a curve class $\mathbf{c}$ with linear system $\mathcal{L}$ has a negative extra-intersection number with a (-1)-curve E, say $\langle \mathcal{L}, E\rangle = -N$ with $N \geq 1$. Since E is an irreducible curve, by Bézout's Theorem every member of the system $\mathcal{L}$ must contain E as a component. We can therefore remove E (subtracting the data of the degree and multiplicity of E from $\mathcal{L}$) and obtain a residual class $\mathcal{L}' = \mathcal{L} - E$. If $N \geq 2$, the residual class also has $\langle \mathcal{L}', E\rangle < 0$, and so E will be removable again. Iterating the analysis implies that a total of N copies of E can be removed from $\mathcal{L}$; algebraically, this means that if the equation of E is $f(x, y) = 0$, then f^N divides the equation of every member of $\mathcal{L}$.

Denote by $\mathcal{M} = \mathcal{L} - NE$ the total residual system, obtained from $\mathcal{L}$ by removing all N copies of E from every member; then $\langle \mathcal{M}, E\rangle = 0$. A straightforward computation using the bilinearity of the extra-intersection number and (3) shows that

$$\nu(\mathcal{M}) = \nu(\mathcal{L}) + N(N-1)/2.$$

Notice that the actual dimensions of the two systems are the same: the elements of $\mathcal{L}$ differ from the elements of $\mathcal{M}$ only by the inclusion of NE as a component, which sets up an isomorphism between the two systems. Therefore, if $N \geq 2$ and the system $\mathcal{M}$ has positive virtual dimension, we see that

$$\begin{aligned}\dim(\mathcal{L}) = \dim(\mathcal{M}) &\geq \nu(\mathcal{M}) \\ &= \nu(\mathcal{L}) + N(N-1)/2 > \nu(\mathcal{L})\end{aligned}$$

and $\mathcal{L}$ is a special system, having dimension strictly greater than its virtual (and hence its expected) dimension.

Example 4. We can reverse this procedure and generate special systems more or less at will by adding in multiple (-1)-curves. Take any nonempty linear system $\mathcal{M}$ and a (-1)-curve E such that the extra-intersection number $\langle \mathcal{M}, E\rangle = 0$. Then the system $\mathcal{L} = \mathcal{M} + NE$ is special for every $N \geq 2$.

The special case when $\mathcal{M}$ is the trivial system gives the examples $\mathcal{L} = NE$, the special systems of multiple (-1)-curves all by themselves.

One can iterate this and continue to add in multiple (-1)-curves as long as one can find disjoint ones (disjoint in the sense of having extra-intersection number zero). A particularly spectacular example is afforded by considering the system $\mathcal{L}_{93}(-57p_0 - \sum_{i=1}^{7} 28p_i)$ of curves of degree 93 with one point of multiplicity 57 and seven points of multiplicity 28. The virtual dimension of this system is $93(96)/2 - 57(58)/2 - 7(28)(29)/2 = -31$, so that this is expected to be quite empty. However, as noted above, through seven general points there is always a cubic that is double at one and passes through the six others; and through eight general points there is always a sextic that is triple at one and passes through the seven others. Consider the seven cubics C_j, $1 \le j \le 7$, with a double point at p_0 and passing through all of the other seven points p_i except for p_j. Let S be the sextic triple at p_0 and double at the other seven p_i's. Then the given system contains as a member the curve $5S + 3\sum_{j=1}^{7} C_j$, and so is not empty at all! These eight curves $S, C_1, \dots, C_7$ are all disjoint (-1)-curves (disjoint in the sense of having extra-intersection number zero).

For those readers who like this sort of thing, one might also consider the linear system $\mathcal{L}_{96}(-\sum_{i=1}^{8} 34p_i)$ of curves of degree 96 with eight points of multiplicity 34. This system has $\nu = -8$, but there is a curve in the system! If one takes the eight sextics which are triple at one of the eight points and double at the other seven, adds them up, and doubles the result, the desired curve emerges.

The Main Conjecture

It is currently the case that the construction described above affords all known examples of special linear systems. Specifically, we have the following formulation:

Main Conjecture. If $\mathcal{L} = \mathcal{L}_d(-\sum m_i p_i)$ is a special linear system for generic points p_i, then there is a (-1)-curve E such that $(\mathcal{L} \cdot E) \le -2$.

More precise versions of the Main Conjecture have been formulated by Hirschowitz in [Hi] and Harbourne in [Ha1] (see also [Ha2]).

Before proceeding to discuss what is known about the conjecture, let us consider some variations to put the problem in a broader context.

Variations and Applications

The Generalization to Higher Dimension

The general problem of computing the dimension of a space of polynomials satisfying certain multiplicity conditions at a set of general points can be formulated in any dimension, not just in the plane. Define $\mathcal{L}_d^{(r)}$ to be the projective space of polynomials (modulo scalars) of degree at most d in r variables; this is a projective space of dimension $\binom{r+d}{d} - 1$. For a polynomial to have multiplicity at least m at a point p, one has $\binom{m-1+r}{r}$ conditions imposed (the number of Taylor coefficients of degree at most $m-1$). If we denote by $\mathcal{L}_d^{(r)}(-\sum_{i=1}^{n} m_i p_i)$ the space of polynomials of degree at most d in r variables having multiplicity at least m_i at n chosen points $\{p_i\}$, the problem is to compute the dimension of this space when the points are chosen generically. As the above discussion indicates, we have a *virtual dimension*

$$\nu = \nu_d^{(r)}\Big(-\sum_{i=1}^{n} m_i p_i\Big) = \binom{r+d}{d} - 1 - \sum_i \binom{m_i - 1 + r}{r}$$

and an *expected dimension* $e = \max\{-1, \nu\}$.

The reader can check that all this generalizes what we have discussed above in the case $r = 2$ and also gives the familiar formulas in the case $r = 1$.

The Alexander-Hirschowitz Theorem

In this general form, the problem of computing the dimension of $\mathcal{L}_d^{(r)}(-\sum_{i=1}^{n} m_i p_i)$ for n general points p_i is unsolved; there is not even a precise conjecture about which of these systems should be "special", in the sense of having dimension higher than expected. However, the analogue of the Multiplicity One Theorem is still true, with practically the same proof.

The only other statement known in higher dimension involves the multiplicity *two* case, which was settled by J. Alexander and A. Hirschowitz about 1988. Here the number of conditions imposed by a point of multiplicity at least two is $r + 1$, so that the expected dimension is

$$\max\{-1, \binom{r+d}{d} - 1 - n(r+1)\}.$$

For a hypersurface to have multiplicity at least two at a point is equivalent to saying that it is singular at the point.

Alexander-Hirschowitz Theorem. Fix $r \ge 2$ and $d \ge 2$, and consider the linear system $\mathcal{L} = \mathcal{L}_d^{(r)}(-\sum_{i=1}^{n} 2p_i)$ consisting of hypersurfaces of degree at most d in r variables that are singular at n general points $\{p_i\}$.

(a) For $d = 2$, the linear system $\mathcal{L}$ is special if and only if $2 \le n \le r$.
(b) For $d \ge 3$, the linear system $\mathcal{L}$ is special if and only if the triple (r, d, n) is one of the following: $(2,4,5), (3,4,9), (4,4,14), (4,3,7)$.

Most of these special systems are rather easily understood. First, let us take up the case of the quadrics, where $d = 2$. Any quadric hypersurface is defined by a quadratic polynomial, which, if we homogenize, can be considered as a quadratic form in $r + 1$ variables. This in turn can be considered as a symmetric square matrix Q of size $r + 1$. We can choose coordinates so that the first $r + 1$ of the points (if there are that many) occur at the "coordinate points" whose homogeneous coordinates correspond to the standard basis vectors; these are the points $[1:0:0:\dots:0]$, $[0:1:0:\dots:0]$, etc. For the quadric hypersurface to have multiplicity at least two at $[1:0:0:\dots:0]$, the first row (and column) of the matrix Q must be zero. This is clearly $r + 1$ linear conditions, as it should be. However, for the quadric to have multiplicity at least two at the second point $[0:1:0:\cdots:0]$, the second row and column of Q must be zero. If the first row and column are already zero, the first entry of the second row and column are automatically zero, so there are only r additional entries that must be zero. Hence the second point imposes only r conditions, not $r+1$, and the actual dimension is one larger than expected. This phenomenon continues until there are $r+1$ points, in which case the matrix Q is all zero, and there are no nontrivial quadratic polynomials satisfying the condition: $\mathcal{L}$ is empty and its actual dimension is -1, which is at this point quite expected.

We saw this example in the plane as the special system $\mathcal{L}_2(-2p-2q)$ of conics double at two points; this exists (as the double line) but is unexpected. Here $d=n=r=2$.

For the special systems with $d=4$, the philosophy is quite similar to that of the (-1)-curves in the plane: in every case the system is expected to be empty, but in fact a double quadric exists. To see the numerology of it all, first note that the space of quadrics in r-space has projective dimension $r(r+3)/2$. Let this be n, so that there exists a unique quadric passing through the n points. Then the double quadric will be a quartic with multiplicity two at the n points. For this to be unexpected, we want $v \le -1$, or

$$\binom{r+4}{4} - 1 - (r(r+3)/2)(r+1) \le -1,$$

which happens exactly for $r=2,3,4$. This explains the first three cases of (b) in the theorem.

The fourth one, the linear systems of cubics in 4-space with multiplicity at least 2 at seven general points, is a different animal altogether. Again the system is expected to be empty. However, there is a cubic with a fascinating construction in the system.

A *rational normal curve* in r-space is a curve that has, after a possible linear change of coordinates, the parametric description

$$t \mapsto (t, t^2, t^3, \ldots, t^r).$$

These curves have a long history and enjoy many properties, causing them to attract more than their share of attention. For example, they have the smallest possible degree (namely r) among all curves spanning r-space. It is an exercise in linear algebra that through $r+3$ general points of r-space there passes a rational normal curve; in particular, through our seven general points of 4-space there passes a rational normal curve C. Let X be the union of all secant lines to C; since C is a curve and a point of a secant line is determined by choosing a first point x on C, choosing a second point y of C, and then choosing a third point on the line $\overline{xy}$, we see that X is a 3-dimensional object in 4-space. It therefore is defined by the vanishing of a single equation F, which is in fact of degree 3: there are exactly three secants to C meeting a general line of 4-space.

This cubic hypersurface X is the unexpected member of the linear system $\mathcal{L}_3^{(4)}(-\sum_1^7 2p_i)$. It clearly passes through the original seven points, since C does; therefore the multiplicity is at least one. If it were only one, say at p_1, then X would be smooth at p_1, not singular. Therefore X would have a tangent space at p_1, which would necessarily have dimension 3 and would contain every secant line of X which contained p_1. But since C spans the 4-space, there are secant lines to C through p_1 in four independent directions! This contradiction shows that X is singular all along C in fact, and in particular at the original seven points.

The reader who is interested in learning a bit more about rational normal curves may consult [Har].

The Waring Problem

The Waring problem for integers is the following: Given an integer N, write it as a sum of d^{th} powers. Of course, one can always do this by using 1^d N times, but the problem begins to have some meat to it when one asks how many d^{th} powers are necessary. For example, the famous Four Squares Theorem says that every positive integer can be written as a sum of four squares; this is sharp, since, for example, 7 is not a sum of three squares.

The Waring problem generalizes to polynomials as follows: Given a homogeneous polynomial ("form") F of degree d, write F as a sum of d^{th} powers of linear forms:

$$F = \sum_i L_i^d.$$

How many powers are necessary for a given F? For every F? For the general F?

It turns out that there is a surprising relationship between Waring's problem for forms and the Alexander-Hirschowitz Theorem on the dimension of linear systems of hypersurfaces with imposed singularities. This relationship exploits the duality between polynomials and partial differential operators.

Since we are discussing forms, it is convenient to work projectively. So fix homogeneous coordinates $[z_0 : z_1 : \cdots : z_r]$ in $\mathbb{P}^r$. Define dual variables $x_0, \ldots, x_r$, that act on the z's as partial derivative operations: $x_i = \partial/\partial z_i$. In this way the dual polynomial ring $\mathbb{C}[x_0, \ldots, x_r]$ acts on the original polynomial ring $\mathbb{C}[z_0, \ldots, z_r]$ so that the homogeneous differential operators in $\underline{x}$ of degree d are perfectly paired with the homogeneous polynomials in $\underline{z}$ of degree d.

Now fix n points $p_1, \ldots, p_n$ in $\mathbb{P}^r$ and a degree d. There are two constructions we can make with these n points. First, we can take the vector space of forms in the z's of degree d that are singular at the p_i's: this is the vector space associated to the space $\mathcal{L}_d^{(r)}(-\sum_i 2p_i)$ discussed above and is the subject of the Alexander-Hirschowitz Theorem.

Second, to any point $q = [q_0 : \cdots : q_r] \in \mathbb{P}^r$ we can associate the linear differential operator $\Delta = \sum_i q_i x_i$, which is well defined up to constant factor. In particular, to a set of n points $\{p_i\}$ in $\mathbb{P}^r$ we obtain a set of n linear differential operators $\Delta_1, \ldots, \Delta_n$ in the dual polynomial ring $\mathbb{C}[x_0, \ldots, x_r]$. Define

$$A_d^{(r)}\left(\sum_i p_i\right) = \left\{\sum_i M_i(\underline{x})\Delta_i^{d-1} \;\middle|\; \deg(M_i) = 1\right\};$$

note that this is a subspace of the space of differential operators of degree d.

Recall that the differential operators in $\underline{x}$ of degree d are perfectly paired with the polynomials in $\underline{z}$ of degree d. In fact, under this pairing, the two spaces $\mathcal{L}_d^{(r)}(-\sum_i 2p_i)$ and $A_d^{(r)}(\sum_i p_i)$ exactly annihilate each other: this is Terracini's Lemma, dating back to 1915 or so.

The proof is not even too hard, once one organizes things and uses all of the algebraic tools at hand. The only real computation to make is to check that it is true for one point, say $p_1 = [1:0:0:\cdots:0]$. Then $\Delta_1 = x_0$, so $A_d^{(r)}(p_1)$ is spanned by the monomials $\{x_i x_0^{d-1} \mid 0 \le i \le r\}$.

Now $\mathcal{L}_d^{(r)}(-2p_1)$ is the space of polynomials of degree d that are singular at p_1, and this implies that such a polynomial cannot contain the monomial z_0^d (else it would not even vanish at p_1) nor any of the monomials $z_i z_0^{d-1}$ (else it would have multiplicity one at p_1).

Visibly these two spaces are dual to each other. This proves the statement for this one special point; it follows for any point by noticing that the duality is equivariant under linear transformations and any two points of $\mathbb{P}^r$ are in the same orbit of $GL(r+1)$.

Now the statement for more than one point follows from the perfection of the pairing if one uses the equalities $A_d^{(r)}(\sum_i p_i) = \sum_i A_d^{(r)}(p_i)$ and $\mathcal{L}_d^{(r)}(-\sum_i 2p_i) = \bigcap_i \mathcal{L}_d^{(r)}(-2p_i)$.

The purpose of noticing this duality is to arrive at a determination of the dimension of $A_d^{(r)}(\sum_i p_i)$:

Corollary. Fix general points $\{p_i\}$. Then

$$\dim A_d^{(r)}\left(\sum_i p_i\right) = \binom{d+r}{r} - 1 - \dim \mathcal{L}_d^{(r)}(-\sum_i 2p_i).$$

If $d \ge 3$, then

$$\dim A_d^{(r)}\left(\sum_i p_i\right) = \min\left\{n(r+1), \binom{d+r}{r}\right\}$$

unless (r, d, n) is one of the four triples $(2,4,5), (3,4,9), (4,4,14), (4,3,7)$, in which case it is one less.

The first statement is simply an application of the duality, and the second uses the Alexander-Hirschowitz Theorem to identify the "unexpected" situations.

Now comes the connection to the Waring problem. Let W be the manifold of all forms of degree d in the x_i's that can be written as a sum of n d^{th} powers of linear forms. Fixing the points p_i as above and the corresponding linear forms Δ_i, we have that

$$A_d^{(r)}\left(\sum_i p_i\right) \text{ is the tangent space to } W \text{ at the point } w = \sum_i \Delta_i^d.$$

This is a simple calculation: the tangent space is given by the first-order terms of the variation

$$\sum_i (\Delta_i + M_i)^d = \sum_i \Delta_i^d + \sum_i d M_i \Delta_i^{d-1} + \ldots,$$

which from the above expansion we see is exactly $A_d^{(r)}(\sum_i p_i)$.

At the general point of W, the dimension of the tangent space is equal to the dimension of W. If we want the general form to be written as a sum of n d^{th} powers, we need W to be the entire space of all forms of degree d, which is equivalent to having $\dim(W) = \binom{d+r}{r}$. Using the above computation, we then need $n(r+1) \ge \binom{d+r}{r}$, unless we are in one of the exceptional cases. This leads to the following.

Waring Problem for General Forms

Fix $d \ge 3$. Then the minimum n such that the general form of degree d in $r+1$ variables can be written as a sum of n d^{th} powers is $\lceil \frac{1}{r+1}\binom{d+r}{r} \rceil$, unless (r, d) is equal to $(2,4)$, $(3,4)$, $(4,4)$, or $(4,3)$, where it requires one more d^{th} power.

The exceptions were known in the last century; Clebsch, Reye, and Sylvester among others had noted them. For the experts, the manifold W introduced above is the "n-secant variety to the d-fold Veronese". Thinking in these terms, Lazarsfeld, Mukai, and others realized the connection between Alexander-Hirschowitz and Waring. The approach above is that taken by J. Emsalem and A. Iarrobino [I].

Recent Work on the Main Conjecture

The Main Conjecture concerning the dimensions of the linear systems of curves in the plane with general multiple base points, i.e., the spaces $\mathcal{L}_d(-\sum_i m_i p_i)$, is still open. Recent work has focused on the special case where all of the multiplicities are equal, say to m; this is the system

$\mathcal{L}_d(-\sum_{i=1}^{n} mp_i)$ of curves of degree d having multiplicity m at n general points. The $m = 2$ case has been treated by B. Segre, by E. Arbarello and M. Cornalba, and by A. Hirschowitz, and is now settled. Hirschowitz also handled the $m = 3$ case in 1985. The $m = 4$ case has recently been addressed by L. Evain and by T. Mignon (who also treats the general case with all $m_i \leq 4$). The Main Conjecture is also true in all of these cases.

In the last ten years there has been a flurry of activity in the area, by those mentioned above, and by L. Caporaso, J. Harris, A. Geramita, A. Gimigliano, Y. Pitteloud, and G. Xu; some of this builds on work of M. Nagata.

Recently the author, in joint work with C. Ciliberto of the University of Rome II, has been investigating a degeneration technique that applies to the problem at hand. The main idea is the following. We are studying systems of curves in the plane $\mathbb{P}^2$, with multiplicity at least m at n general points. If we put the points in special position, the dimension of the space can only rise, by semicontinuity. On the other hand, the special position of the points may allow us to compute the dimension more easily. If we are able to find a special position for the points such that the resulting system has the expected dimension, then it will certainly have the expected dimension for general positions of the points.

This "degeneration" approach, namely, to degenerate the positions of the points, is fundamentally the approach of the previous authors, in one form or another. The problem is that the computation of the dimension for the case when the points are in special position is, by its very nature, a somewhat special computation, and therefore it is difficult to arrange the analysis to take advantage of any possible recursions that may present themselves.

Ciliberto and the author have studied the possibility of degenerating the entire plane itself to two surfaces, each of which is more or less a plane. The corresponding degeneration in one dimension can be described as follows. Take the line with affine coordinate t. Embed the line in the plane using the coordinatization of the rational normal curve of degree 2: $t \mapsto (t, t^2)$. This realizes the line as a conic (a parabola, in fact) in the (x, y) plane, given by $y = x^2$. Now degenerate the conic to two lines by introducing a degeneration parameter u and considering $(1 - u)y^2 + uy = x^2$. For $u = 1$ we have the original parabola; for u near zero but not zero, we have hyperbolas; for $u = 0$ we have the two lines $y = \pm x$.

This trick, of degenerating one line to two, can be executed with the plane also using similar methods. One re-embeds the plane into 5-space (by sending (x, y) to (x, y, x^2, xy, y^2)); the result is a quartic surface, which then degenerates to a union of two surfaces, one a plane and one a cubic surface that is itself a re-embedded plane. This has the effect of degenerating the plane into two planes. A degeneration of this type was first exploited by Z. Ran for enumerative purposes in [R].

One now studies the degeneration of the plane curves, passing through n general points with multiplicity at least m. The advantage of this approach is that the n points also degenerate to points on the two surfaces, and one can arrange to degenerate a of the points on the one plane and $b = n - a$ of the points on the other plane. The plane curves then degenerate to plane curves of known degree having multiplicity at least m at a points (respectively b points) in the limit. The reader can now appreciate the possibility of a recursion being performed, which is in fact what we are able to execute in many cases.

The story gets technical rather quickly, but the bottom line is that using this method we have been able to verify that the Main Conjecture is true whenever the multiplicities are all constant and at most 12; these results are described in [CM1] and [CM2]. Unfortunately, we have found examples of parameters d, m, and n for which not one of these degenerations works, in the sense that the dimension of the limit system is always greater than the expected dimension: hence no conclusion can be drawn from the semicontinuity argument. So more thought is required!

I hope that I have piqued the reader's interest in some of these questions and given a small glimpse into the theoretical world of polynomial interpolation and its ramifications. In summary, here is a great classical problem, easily stated and understood, which is stubbornly resisting attempts to finish it off.

Acknowledgments

This article grew out of two sets of lectures given by the author in 1997, one at the University of Rome II in June and the other at the University of Bayreuth in October. I wish to thank C. Ciliberto and C. DeConcini for their hospitality in Rome and T. Peternell and F. Schreyer for theirs in Bayreuth. I especially want to acknowledge Michael Schneider, who invited and encouraged me to give the summer school lectures in Bayreuth and who tragically passed away before I had a chance to see him there. He is greatly missed by a host of colleagues around the globe who benefited from his energy, enthusiasm, expertise, and excitement for mathematics and for life.

References

[CM1] C. Ciliberto and R. Miranda, Degenerations of planar linear systems, *J. Reine Angew. Math.* **501** (1998), 191–220.

[CM2] ____ , Linear systems of plane curves with base points of equal multiplicity, to appear in *Trans. Amer. Math. Soc.*

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

[F] W. Fulton, *Algebraic Curves*, Math. Lecture Note Ser., Benjamin, New York, 1969.

[Ha1] B. Harbourne, The geometry of rational surfaces and Hilbert functions of points in the plane, *Proceedings of the 1984 Vancouver Conference in Algebraic Geometry*, CMS COnf. Proc., vol. 6, Amer. Math. Soc., Providence, RI, 1986, pp. 95–111.

[Ha2] ____ , Points in good position in $\mathbb{P}^2$, *Zero-Dimensional Schemes (Ravello 1992)*, de Gruyter, Berlin, 1994, pp. 213–229.

[Har] J. Harris, *Algebraic Geometry: A First Course*, Graduate Texts in Math., vol. 133, Springer-Verlag, Berlin and New York, 1992.

[Hi] A. Hirschowitz, Une conjecture pour la cohomologie des diviseurs sur les surfaces rationnelles génériques, *J. Reine Angew. Math.* **397** (1989), 208–213.

[I] A. Iarrobino, The inverse system of a symbolic power, II: The Waring problem for forms, *J. Algebra* **174** (1995), 1091–1110.

[R] Z. Ran, Enumerative geometry of singular plane curves, *Invent. Math.* **97** (1989), 447–465.