Polynomials

A chapter for the Mathematics++ Lecture Notes

Jiří Matoušek

Rev. 19/V/14 JM

Here we discuss polynomials in several variables. They belong among the most powerful and most often applied mathematical tools in computer science, and sometimes their use works like a magic wand.

The set of all solutions of a system of m polynomial equations in n variables is called an *algebraic variety*, and it is studied in *algebraic geometry*, one of the most classical and deepest areas of mathematics. Here we will make the first few steps in this fascinating field.

1 Rings, fields, and polynomials

A *ring* R is an algebraic structure with addition and multiplication; the readers unsure about the definition might want to check it. Here we will consider only *commutative* rings (commutativity concerns multiplication, since addition in a ring is always commutative) with *unit element* 1. Actually, unlike in usual introductory courses of algebra, we will see a large menagerie of rings.

A *field* is a ring in which we also have division (by each nonzero element, that is). We will most often consider the fields $\mathbb{R}$, the reals, and $\mathbb{C}$, the complex numbers, sometimes also a finite field $\mathbb{F}_q$ with q elements, where q, as we recall, must be a prime power, and the rationals $\mathbb{Q}$. An arbitrary field will usually be denoted by $\Bbbk$, partially in agreement with a typical convention in algebraic geometry where they use k.

Everyone knows univariate polynomials such as $37x^5 - 2x^4 + 12$. The set of all polynomials in a variable x with coefficients in a ring R is denoted by $R[x]$. It also forms a ring, with the usual addition and multiplication of polynomials.

We will more often consider *multivariate* polynomials, such as $13x^5y^3z - 6x^2y^4z^2 + y^2 - 2z$. We write $R[x_1, \ldots, x_n]$ for the ring of all polynomials in variables $x_1, \ldots, x_n$ with coefficients in R. A polynomial $f \in R[x_1, \ldots, x_n]$ is a finite sum of **terms** of the form $c_\alpha x^\alpha$, where $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$ is a vector of nonnegative integers, $c_\alpha \in R$ is a **coefficient**, and x^α is a convenient shorthand for the **monomial** $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. The **degree** of such a monomial is $\alpha_1 + \cdots + \alpha_n$. The degree of f, written $\deg f$, is the maximum of the degrees of its monomials.

The degree of the zero polynomial, which has no monomials, is usually taken as $-\infty$.

Each polynomial $f \in R[x_1, \ldots, x_n]$ defines a function $R^n \to R$ in an obvious way. Usually we will use the same letter for the polynomial and for the function.

Exercise 1.1. *Prove (before reading further, and carefully!) that if the function defined by a polynomial $f \in \mathbb{R}[x, y]$ is zero everywhere on $\mathbb{R}^2$, then f is the* **zero polynomial***; that is, all coefficients are* 0*. Similarly for $\Bbbk[x_1, \ldots, x_n]$ where $\Bbbk$ is an* infinite *field. (Note that if $\mathbb{F}$ is a finite field, then $\prod_{a\in\mathbb{F}}(x-a)$ is a nonzero polynomial defining the zero function $\mathbb{F} \to \mathbb{F}$.)*

Rigidity of polynomials. Polynomials constitute one of the most significant classes of functions, and they have various amazing properties. For computer science, one of the most important properties is some kind of *rigidity*, which can be vaguely expressed as "if two polynomials differ, then they differ *a lot*."

Here is a well-known manifestation of rigidity for univariate polynomials.

(Univariate rigidity) A nonzero univariate polynomial $f \in \Bbbk[x]$ of degree $d \geq 0$, where $\Bbbk$ is a *field*, has at most d roots. Consequently, if $f, g \in \Bbbk[x]$, $\deg(f), \deg(g) \leq d$, and $f(a) = g(a)$ for at least $d + 1$ distinct points a, then $f = g$.

We recall that this is proved by induction on d, by checking that if $f(a) = 0$, then f is divisible by $x - a$.

2 The Schwartz–Zippel theorem

This is a manifestation of rigidity in the multivariate case, one which is quite simple to prove and extremely useful, e.g., for randomized algorithms.

Theorem 2.1 (The Schwartz–Zippel theorem)**.** *Let $\Bbbk$ be a field, let d be a natural number, and let S be a finite subset of $\Bbbk$. Then for every nonzero polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ of degree d, the number of n-tuples $(r_1, r_2, \ldots, r_n) \in S^n$ with $f(r_1, \ldots, r_n) = 0$ is at most $d|S|^{n-1}$. In other words, if $r_1, \ldots, r_n \in S$ are chosen independently and uniformly at random, then the probability of $f(r_1, \ldots, r_n) = 0$ is at most $\frac{d}{|S|}$.*

Here we measure the size of the zero set of f discretely, by counting the points of its intersection with the "combinatorial cube" S^n. If $\Bbbk = \mathbb{F}_q$, we can often simply take $S = \mathbb{F}_q$.

Proof of the Schwartz–Zippel theorem. We proceed by induction on n. The $n = 1$ case was mentioned earlier, so let $n > 1$.

Let us suppose that x_1 occurs in at least one term of f with a nonzero coefficient (if not, we rename the variables). Let us write f as a polynomial in x_1 with coefficients being polynomials in $x_2, \ldots, x_n$:

$$f = \sum_{i=0}^{k} f_i x_1^i, \quad f_1, \ldots, f_k \in \Bbbk[x_2, \ldots, x_n],$$

where k is the maximum exponent of x_1 in f.

We divide the n-tuples $(r_1, \ldots, r_n)$ with $f(r_1, \ldots, r_n) = 0$ into two classes. The first class, called V_1, consists of the n-tuples with $f_k(r_2, \ldots, r_n) = 0$. Since the polynomial $f_k(x_2, \ldots, x_n)$ is not identically zero and has degree at most $d - k$, the number of choices for $(r_2, \ldots, r_n)$ is at most $(d-k)|S|^{n-2}$ by the induction hypothesis, and so $|V_1| \leq (d-k)|S|^{n-1}$.

The second class V_2 are the remaining n-tuples, that is, those with $f(r_1, \ldots, r_n) = 0$ but $f_k(r_2, \ldots, r_n) \neq 0$. Here we count as follows: r_2 through r_n can be chosen in at most $|S|^{n-1}$ ways, and if $r_2, \ldots, r_n$ are fixed with $f_k(r_2, \ldots, r_n) \neq 0$, then r_1 must be a root of the univariate polynomial $g(x_1) = f(x_1, r_2, \ldots, r_n)$. This polynomial has degree (exactly) k, and hence at most k roots. Thus $|V_2| \leq k|S|^{n-1}$, which gives $d|S|^{n-1}$ altogether, finishing the proof. □

Exercise 2.2. *Check that the Schwartz–Zippel theorem is tight; i.e., exhibit an n-variate polynomial of degree d whose zero set in S^n has $d|S|^{n-1}$ points (where $d < |S|$).*

A well known "continuous" counterpart of the Schwartz–Zippel theorem asserts that the zero set of a nonzero polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ is a Lebesgue null set.[1] This follows, e.g., from *Sard's theorem* of mathematical analysis, or it can be proved directly.

Exercise 2.3. *Imitate the proof of the Schwartz–Zippel theorem to show that the zero set of f is Lebesgue null for every nonzero $f \in \mathbb{R}[x_1, \ldots, x_n]$. (Fubini's theorem heps with a convenient proof, if you are somewhat familiar with measure theory.)*

3 Polynomial identity testing

Testing perfect matchings. We recall that a **matching** in a graph G is a set of edges $F \subseteq E(G)$ such that no vertex of G is incident to more than one edge of F. A **perfect matching** is a matching covering all vertices. One of the most famous uses of the Schwartz–Zippel theorem is in an algebraic algorithm for testing the existence of a perfect matching in a given graph.

For simplicity, we will discuss only the bipartite case. So we consider a bipartite graph, with vertices divided into two classes $\{u_1, u_2, \ldots, u_n\}$ and $\{v_1, v_2, \ldots, v_n\}$ and the edges going only between the two classes. Both of the classes have the same size, for otherwise, there is no perfect matching. Let m stand for the number of edges of G.

Let S_n be the set of all permutations of the set $\{1, 2, \ldots, n\}$. Every perfect matching F of G uniquely corresponds to a permutation $\pi \in S_n$; we can write $F = \{\{u_1, v_{\pi(1)}\}, \{u_2, v_{\pi(2)}\}, \ldots, \{u_n, v_{\pi(n)}\}\}$.

We express the existence of a perfect matching by a determinant, but not of an ordinary matrix of numbers, but rather of a matrix whose entries are

[1] We recall that a set $X \subseteq \mathbb{R}^n$ is *Lebesgue null* if, for every $\varepsilon > 0$, X can be covered by at most countably many axis-parallel boxes of total volume at most ε. Instead of Lebesgue null, one also says that X has (Lebesgue) measure zero.

variables. We introduce a variable x_{ij} for every edge $\{u_i, v_j\} \in E(G)$ (so we have m variables altogether), and we define an $n \times n$ matrix A by

$$a_{ij} := \begin{cases} x_{ij} & \text{if } \{u_i, v_j\} \in E(G), \\ 0 & \text{otherwise.} \end{cases}$$

The determinant of A is a polynomial in the m variables x_{ij}. By the definition of the determinant,

$$\begin{aligned} \det A &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \cdot a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)} \\ &= \sum_{\substack{\pi \text{ describes a perfect} \\ \text{matching of } G}} \operatorname{sgn}(\pi) \cdot x_{1,\pi(1)} x_{2,\pi(2)} \cdots x_{n,\pi(n)}. \end{aligned}$$

Clearly, if G has no perfect matching, then $\det A$ is the zero polynomial. But the converse also holds: if G does have a perfect matching, then $\det A \neq 0$ as a polynomial. To see this, we fix a permutation σ that defines a perfect matching, and we set $x_{i,\sigma(i)} := 1$ for every $i = 1, 2, \ldots, n$, while the remaining x_{ij} are set to 0. Then all terms in the above expansion of $\det A$ vanish *except* for the one corresponding to σ, which is ± 1.

So testing for a perfect matching in G is equivalent to testing if $\det A$ is the zero polynomial. We cannot afford to compute $\det A$ explicitly in the usual form, as a sum of monomials, since it may have up to $n!$ terms.

But if we substitute specific numbers for the variables x_{ij}, we can calculate the determinant reasonably fast, e.g., by Gaussian elimination. So we can imagine that $\det A$ is available through a black box, from which we can obtain its value at any specified point.

Since $\deg(\det A) \leq n$, the Schwartz–Zippel theorem shows that if $\det A$ is nonzero and we compute it for values of the x_{ij} chosen independently at random from $S := \{1, 2, \ldots, 2n\}$, then the probability of getting 0 is at most $\frac{1}{2}$. This gives a probabilistic algorithm for testing the existence of a perfect matching in G.

The probability of error can be reduced, either by repeating the algorithm several times, or by choosing from a larger set S.

Computationally, instead of working over the integers, it is better to compute the determinant over a sufficiently large finite field, because then we need not worry about the intermediate values in the computation getting very large. (There *is* a polynomial-time version of the Gaussian elimination over the integers, but it is not an easy matter.)

If we compute the determinant by Gaussian elimination, the running time is $O(n^3)$, which is worse than for some combinatorial algorithms for perfect matchings. But using fast matrix multiplication, the determinant can be computed faster; the current best asymptotic running time is $O(n^{2.376})$. This yields the asymptotically fastest known perfect matching algorithm.

This algorithm also has a fast implementation on a parallel computer, with polylogarithmic running time. No other known approach yields comparably fast parallel algorithms.

Finally, it is worth mentioning that although the basic version of the algorithm, as described above, only decides if there is a perfect matching but does not find one, there are more sophisticated extensions that also find a perfect matching, and if a perfect matching does not exist, they can find a matching of maximum cardinality. See [Har09] for recent results and references.

Counting compositions. The strategy in the above algebraic algorithm is very general and can be used for an arbitrary *polynomial identity testing*; that is, for a polynomial of controlled degree provided by a black box, the Schwartz–Zippel theorem allows us to test whether the polynomial is identically zero. Here is another lovely application.

Given a set $P \subseteq S_n$ of permutations, we would like to count $|P \circ P|$, i.e., the number of distinct permutations ρ that can be expressed as a composition $\sigma \circ \tau$ for $\sigma, \tau \in P$. Mainly for notational simplicity, let us assume $|P| = n$.

A straightforward algorithm for computing $|P \circ P|$ takes every pair $(\sigma, \tau) \in P^2$, computes the composition $\sigma \circ \tau$ in $O(n)$ time, and then counts the number of distinct permutations in the resulting list. With some care, this can be implemented in a total of $O(n^3)$ time.

To get an asymptotically faster, algebraic algorithm, we introduce variables $x_1, \dots, x_n$ and $y_1, \dots, y_n$. Let us observe that, given permutations σ and τ, the (quadratic) polynomial

$$f_{\sigma\tau} := \sum_{i=1}^{n} x_{\sigma(i)} y_{\tau^{-1}(i)}$$

encodes the composition $\rho := \sigma \circ \tau$, in the sense that

$$f_{\sigma\tau} = \sum_{i=1}^{n} x_{\rho(i)} y_i,$$

as is easy to check. Consequently, $f_{\sigma\tau}$ and $f_{\sigma'\tau'}$ are equal polynomials iff $\sigma \circ \tau = \sigma' \circ \tau'$. Hence, $|P \circ P|$ equals the number of distinct polynomials among $(f_{\sigma\tau} : \sigma, \tau \in P)$.

Next, we observe that all of the $f_{\sigma\tau}$ can be evaluated simultaneously using a matrix product. Indeed, let us enumerate $P = \{\sigma_1, \dots, \sigma_n\}$, and define the polynomial matrices A, B with $a_{ij} = x_{\sigma_j(i)}$ and $b_{ij} = y_{\sigma_j^{-1}(i)}$. Setting $C = A^T B$, we find that $c_{ij} = f_{\sigma_i\sigma_j}$.

The probabilistic algorithm for computing $|P \circ P|$ now goes as follows. We set $N := 4n^4$, $S := \{1, 2, \dots, N\}$, we pick values $s_1, \dots, s_n$ and $t_1, \dots, t_n$ from S independently and uniformly at random, we make the substitutions $x_i := s_i$ and $y_i := t_i$, $i = 1, 2, \dots, n$, and we compute the value of C. By fast matrix multiplication this can be done in $O(n^{2.376})$ time. We return the number of distinct entries of the resulting matrix as the answer.

Clearly, this answer is never larger than $|P \circ P|$. If it is strictly smaller than $|P \circ P|$, it means that a nonzero polynomial of the form $f_{\sigma\tau} - f_{\sigma'\tau'}$ evaluates to 0 at $s_1, \dots, s_n$, $t_1, \dots, t_n$. For every fixed fourtuple $(\sigma, \tau, \sigma', \tau') \in P^4$, this has probability at most $\frac{1}{2n^4}$ according to the Schwartz–Zippel theorem (with degree $d = 2$). The probability that this occurs for at least one of the at most n^4 fourtuples is thus at most $\frac{1}{2}$.

Hence the answer is correct with probability at least $\frac{1}{2}$. As before, this probability can be boosted by repetition and/or by choosing larger S.

4 Interpolation, joints, and contagious vanishing

We begin with a small counting question, whose result appears very often when dealing with polynomials.

Fact 4.1. *The number of monomials of degree at most d in variables $x_1, \ldots, x_n$ equals $\binom{d+n}{n}$.*

Indeed, the number in question is the number of ordered n-tuples $(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$ with $\alpha_1 + \cdots + \alpha_n \leq d$, and counting them is basic combinatorics which we omit here.

Somewhat imprecisely, we can say that $\binom{d+n}{n}$ is the number of "degrees of freedom" for a general n-variate polynomial of degree at most d. In a more sophisticated language, if $P_d \subset \Bbbk[x_1, \ldots, x_n]$ is the vector space (over $\Bbbk$) consisting of all polynomials of degree at most d, then all monomials of degree at most d form a basis, and so $\dim P_d = \binom{d+n}{n}$ by Fact 4.1.

The next simple but surprisingly useful lemma can be regarded as a kind of counterpart of the Schwartz–Zippel theorem: that theorem says that the zero set of a polynomial cannot be too big, and the lemma tells us that, nevertheless, we can cover a significant number of points by such a zero set.

Lemma 4.2. *Let $a_1, a_2, \ldots, a_N$ be points in $\Bbbk^n$, where $N < \binom{d+n}{n}$. Then there exists a nonzero polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ of degree at most d such that $f(a_i) = 0$ for all i.*

Proof. Given the a_i, we regard the coefficients c_α of the desired polynomial f as unknowns. So we have $\binom{d+n}{n}$ unknowns. A requirement of the form $f(a) = 0$ translates to a *homogeneous* linear equation for the c_α. Since $N < \binom{n+d}{n}$, we have fewer equations than unknowns, and such a homogeneous system always has a nonzero solution. So there is a polynomial with at least one nonzero coefficient.

Expressed differently, we can consider the linear map $P_d \to \Bbbk^N$ that sends a polynomial f to the N-tuple $(f(a_1), \ldots, f(a_N))$. Since $\dim P_d > N$, this map has a nontrivial kernel. □

Exercise 4.3. (a) *We recall that real numbers $\xi_1, \ldots, \xi_n$ are* **algebraically independent** *(over the rationals) if there is no nonzero polynomial $f \in \mathbb{Q}[x_1, \ldots, x_n]$ with $f(\xi_1, \ldots, \xi_n) = 0$. Prove that for every n there exist n algebraically independent real numbers. Hint: one can use a cardinality argument or a measure argument, for example.*

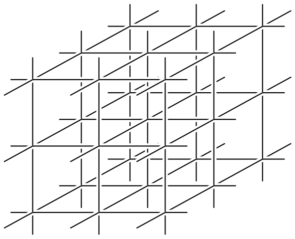
(b) *Show that if $a_1, \ldots, a_N \in \mathbb{R}^n$ are points whose nN coordinates are algebraically independent, and if $N = \binom{d+n}{n}$, then the only polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ of degree at most d vanishing at all the a_i is identically zero.*

Exercise 4.4. (a) *Given $a_1, \ldots, a_N \in \Bbbk^n$ and values $b_1, \ldots, b_N \in \Bbbk$, prove that there exists a polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ with $f(a_i) = b_i$ for all $i = 1, \ldots, N$, and with* $\deg f \leq N - 1$.

(b) *Show that the bound* $\deg f \leq N - 1$ *is optimal in the worst case (i.e., find $a_1, \ldots, a_N$ and $b_1, \ldots, b_N$ for which no f of smaller degree will do). Note that for $n \geq 2$, this bound is very different from the one in Lemma 4.2.*

The joints problem. We consider a set L of n lines in $\mathbb{R}^3$, and call a point $a \in \mathbb{R}^3$ a *joint* if there are at least three lines of L, not all coplanar, passing through a. The question is, what is the maximum possible number of joints for n lines?

There is a lower bound of $\Omega(n^{3/2})$ attained by a grid of lines,



and it was proved by Guth and Katz [GK10] in 2008, after many years of effort by a number of people and many intermediate results, that, asymptotically, this is the most one can get.

Theorem 4.5. *The maximum number of joints of n lines in $\mathbb{R}^3$ is $O(n^{3/2})$.*

There is a straightforward generalization to $\mathbb{R}^d$: for every fixed d, the maximum number of joints of n lines in $\mathbb{R}^d$ is of order $n^{d/(d-1)}$, where a joint means a point common to at least d lines whose direction vectors span $\mathbb{R}^d$. For simplicity we stick to the $d = 3$ case.

On partial derivatives. We recall that, for a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$, the **partial derivative** $\partial f/\partial x_i$ is the usual derivative of a univariate real function, where x_i is regarded as a variable, while all the other x_j are considered constant. The **gradient** of f is the n-tuple

$$\nabla f := \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right).$$

As a side remark, we note that the derivative can be defined purely formally, by setting $\partial(x^i)/\partial x := ix^{i-1}$ and extending linearly, and this makes sense over any field. Many of the usual properties of derivatives can then be checked as well, and so one need not specialize to real (or complex) numbers. For finite fields, though, it may be better to work with the *Hasse derivative*, where the mth Hasse derivative $D^m(x^i) := \binom{i}{m}x^{i-m}$; this avoids troubles with dividing by zero, e.g., in a Taylor expansion formula.

If $f \in \mathbb{R}[x_1, \ldots, x_n]$ is of degree $d \geq 1$, then each of the partial derivatives $\partial f/\partial x_i$ is a polynomial of degree at most $d - 1$, and at least one of them is nonzero.

The following observation connects the definition of a joint to algebra.

Observation 4.6. *Let a be a joint of lines ℓ_1, ℓ_2, ℓ_3 in $\mathbb{R}^3$, and let $f \in \mathbb{R}[x_1, x_2, x_3]$ be a polynomial that vanishes on each of the ℓ_i. Then $\nabla f(a) = 0$; that is, all of the partial derivatives of f vanish at a.*

Proof. This follows easily using the notion and simple properties of directional derivatives. Here is a more explicit argument.

W.l.o.g. we may assume $a = 0$ (the general case follows by translation). If we write $f = c_0 + c_1x_1 + c_2x_2 + c_3x_3 + \textit{terms of degree} \geq 2$, then we have $\frac{\partial f}{\partial x_i}(0) = c_i$, and $\nabla f(0) = c := (c_1, c_2, c_3)$. Letting $v_i = (v_{i1}, v_{i2}, v_{i3})$ be the directional vector of ℓ_i, the restriction of f to the line ℓ_i can be regarded as the univariate polynomial

$$f(tv_i) = c_0 + (c_1v_{i1} + c_2v_{i2} + c_3v_{i3})t + (\textit{terms of degree} \geq 2).$$

Thus, vanishing of this polynomial means, in particular, that $c_1v_{i1} + c_2v_{i2} + c_3v_{i3} = 0$; that is, the vector c is perpendicular to v_i. A vector perpendicular to three linearly independent vectors in $\mathbb{R}^3$ must be zero. □

We are almost ready for the proof of the $O(n^{3/2})$ upper bound for joints, but there is still a simple technical step to be prepared. If we have a set L of n lines with a large number m of joints, then an average line contains "many" joints, namely, $3m/n$. But for the polynomial argument to work, we want that *every* line contains many joints. This is taken care by a standard "pruning" argument (if you know the proof of the statement that every graph of average degree 2δ contains a subgraph with minimum degree at least δ, then you also know the proof of the next lemma).

Lemma 4.7. *Let L be a set of n lines in $\mathbb{R}^3$, let J be the set of all joints of L, and let $b := m/2n$, where $m = |J|$. Then there are subsets $L' \subseteq L$ and $J' \subseteq J$ such that $L' \neq \emptyset$, every point of J' is a joint of the lines of L', and every line of L' contains more than b points of J'.*

Proof. We use the following *pruning* procedure: We set $J_0 = J$, $L_0 = L$, and for $i = 1, 2, 3, \ldots$, if L_{i-1} contains a line ℓ with at most b joints of J_{i-1}, we set $L_i := L_{i-1} \setminus \{\ell\}$, and $J_i := J_{i-1} \setminus \ell$ (i.e., all joints in which ℓ participated are removed).

By definition, this procedure finishes with some $L' = L_k$ and $J' = J_k$ such that each point of J' is a joint of the lines in L' and each line of L' contains more than b joints of J'.

It remains to verify that $L' \neq \emptyset$, for which it suffices to check $J' \neq \emptyset$. Since we have removed at most n lines and at most b joints per line, we have $|J'| \geq m - nb = m/2 > 0$. □

Now we can focus on the essence.

Proof of Theorem 4.5. For contradiction, we suppose that a set L of n lines has $m \geq 7n^{3/2}$ joints. Let J, $b = m/2n$, J', and L' be as in the previous lemma.

We choose a nonzero polynomial $f \in \mathbb{R}[x_1, x_2, x_3]$ that vanishes on all of J' and, subject to this condition, has the *smallest possible degree.*

First we claim that $\deg f \le b$. Indeed, by Lemma 4.2, $\deg f$ does not exceed the smallest integer d with $\binom{d+3}{3} > |J'|$, and a simple calculation shows that $\binom{b+3}{3} > m \ge |J'|$. Namely,

$$\binom{b+3}{3} > \frac{b^3}{3!} = \frac{(m/2n)^3}{3!} = m\frac{m^2}{48n^3} \ge m,$$

since we assumed $m \ge 7n^{3/2}$.

The restriction of f on each line $\ell \in L'$ is thus a univariate polynomial of degree at most $\deg f \le b$ that vanishes in at least $b+1$ points, and hence f vanishes everywhere on ℓ. By Observation 4.6, we have all the partial derivatives $\partial f/\partial x_i$ zero on all of J'. At the same time, since $\deg f \ge 1$, at least one of these partial derivatives is a nonzero polynomial.

But then such a nonzero partial derivative is a polynomial of degree strictly smaller than f vanishing on J', and this contradicts our choice of f and concludes the proof. □

This kind of argument is what Larry Guth calls *contagious vanishing*: the vanishing of f spreads like infection from J' to the lines of L'. In more complicated proofs of this kind, this spreading may continue further, to suitable planes or surfaces, and sometimes ultimately to the whole space.

There are several other beautiful applications of the contagious vanishing argument. The most significant ones are probably a near-optimal solution to the *Erdős distinct distances problem* due to Elekes, Guth, and Katz, and a proof of the *Kakeya conjecture over finite fields* due to Dvir. These, and much more, can be found, e.g., in Guth's book [Gut13] in preparation or in Tao's survey [Tao13]. There are also older arguments in number theory, due to Thue (see [Gut13]) and, especially, due to Baker (see [Wal79, Sec. 4]), which use some sort of contagious vanishing.

5 Varieties, ideals, and the Hilbert basis theorem

Varieties. Let $\mathcal{F} \subseteq \Bbbk[x_1,\ldots,x_n]$ be a set of polynomials, possibly infinite. The **variety** $V(\mathcal{F})$ of $\mathcal{F}$ is the set of common zeros of the polynomials in $\mathcal{F}$:

$$\boxed{V(\mathcal{F}) := \{(a_1,\ldots,a_n) \in \Bbbk^n : f(a_1,\ldots,a_n) = 0 \text{ for all } f \in \mathcal{F}\}.}$$

Some sources use Z instead of V, Z for "zero set."

An **(algebraic) variety**[2] is any subset of $\Bbbk^n$ that can be expressed as $V(\mathcal{F})$ for some $\mathcal{F}$. More precisely, such a set is called an **affine** algebraic variety, to distinguish it from a *projective* algebraic variety, to be mentioned later.

Exercise 5.1. *Show that a finite union, as well as an arbitrary intersection, of varieties is a variety.*

[2] In some of the sources an algebraic variety is also required to be irreducible (this is a notion defined later), while an arbitrary $V(\mathcal{F})$ is called an *algebraic set*.

Exercise 5.2. *Prove that the sets $\mathbb{Z} \subseteq \mathbb{R}$ and $[0,1]^2 \subseteq \mathbb{R}^2$ are not algebraic varieties (over $\mathbb{R}$).*

Algebraic geometry. The study of algebraic varieties is called *algebraic geometry.* In the literature, one can encounter (at least) two quite distinct branches of algebraic geometry, with different flavor and conventions.

Classical algebraic geometry mainly investigates varieties over the complex numbers and, more generally, over **algebraically closed** fields (these are fields in which every nonconstant polynomial has a root). It is an enormous, very important, highly developed, and sometimes very abstract area of modern mathematics. Actually, since the work of Grothendieck in the 1960s, "true" algebraic geometers no longer consider algebraic varieties, but rather *schemes.* A scheme is a more general and technically convenient notion, for which even the definition is out of our scope; see, e.g., [Gat13] for an introduction.

Real algebraic geometry considers varieties over $\mathbb{R}$ and, more generally, **semialgebraic sets**, which are defined not only by conjunctions of polynomial equations, but also by Boolean combinations of polynomial inequalities. One can say that the results are perhaps less elegant than those about varieties over algebraically closed fields, but sometimes they are closer to the needs of computer science and other applications.

We will see a sample of basic results in both of these directions.

Ideals. We recall that an **ideal** in a (commutative) ring R is a subset $I \subseteq R$ that contains 0 and is closed under addition and under multiplication by arbitrary elements of R (in symbols, $f, g \in I$ implies $f + g \in I$ and $f \in I$, $h \in R$ implies $hf \in I$).

Exercise 5.3. *Show that a ring R (commutative and with 1) has only two ideals, $\{0\}$ and R, iff it is a field.*

For a subset $\mathcal{F} \subseteq R$, we let $\langle \mathcal{F} \rangle$ be the **ideal generated by $\boldsymbol{\mathcal{F}}$**. By definition, this is the intersection of all ideals in R that contain $\mathcal{F}$, and it is easy to check that $\langle \mathcal{F} \rangle = \{h_1 f_1 + \cdots + h_n f_n : n \geq 0, f_1, \ldots, f_n \in \mathcal{F}, h_1, \ldots, h_n \in R\}$ (this is similar to linear combinations in linear algebra, but here we multiply by arbitrary elements of R).

Specializing this to the polynomial ring $\mathbb{k}[x_1, \ldots, x_n]$, it is easy to see that for every set $\mathcal{F}$ of polynomials we have $V(\mathcal{F}) = V(\langle \mathcal{F} \rangle)$. Therefore, every variety X is the set of common zeros of some ideal I in $\mathbb{k}[x_1, \ldots, x_n]$; $X = V(I)$. Ideals are usually much better to work with than arbitrary sets of polynomials.

Here is the first significant general result about varieties: we can restrict ourselves to *finitely generated* ideals. A ring R is called **Noetherian** if every ideal in R is generated by a finite set. In particular, every field $\mathbb{k}$ is Noetherian, since the only ideals are $\{0\}$ and $\mathbb{k} = \langle 1 \rangle$.

In the literature, the definition is often stated in a different but equivalent way: R is Noetherian iff there is no infinte sequence of properly nested ideals $I_1 \subsetneq I_2 \subsetneq \cdots$ in R.

Exercise 5.4. *Check this equivalence.*

Theorem 5.5 (Hilbert basis[3] theorem)**.** *If R is a Noetherian ring, then the polynomial ring $R[x]$ is Noetherian as well. Consequently, $\Bbbk[x_1, \ldots, x_n]$ is Noetherian for every field $\Bbbk$.*

Hilbert's proof, more than 100 years old, was unusual at that time since it was nonconstructive: it proved the existence of a finite generating set in every ideal of $\Bbbk[x_1, \ldots, x_n]$, but did not provide any method for finding one. This nonconstructive approach was initially criticized, but later on embraced enthusiastically by the mathematical community. In the last decades, with renewed emphasis on computations and algorithms, people again put much effort into finding constructive, and efficient, proofs for important results.

Proof. Let $I \subset R[x]$ be an ideal. We are going to choose a sequence $f_1, f_2, f_3, \ldots$ of elements (polynomials) from I inductively: for $i = 1, 2, 3, \ldots$, f_i is an element of the smallest degree in $I \setminus \langle f_1, \ldots, f_{i-1} \rangle$. For $i = 1$, in particular, we have $\langle \emptyset \rangle = \{0\}$, and so f_1 is a smallest-degree nonzero element of I. We have $\deg f_1 \le \deg f_2 \le \cdots$.

If we reach some n with $\langle f_1, \ldots, f_n \rangle = I$, we are done.

Otherwise, let $a_i \in R$ be the **leading coefficient** of f_i (the coefficient of the highest power of x), and let us consider the ideal $L = \langle a_1, a_2, \ldots \rangle$. Since R is Noetherian, L is generated by $a_1, \ldots, a_m$ for some finite m.

We claim that $I' := \langle f_1, \ldots, f_m \rangle = I$. If not, then f_{m+1} was chosen as a smallest-degree element in $I \setminus I'$. The leading coefficient a_{m+1} of f_{m+1} belongs to L and thus it can be written as $a_{m+1} = \sum_{i=1}^{m} h_i a_i$, where $h_1, \ldots, h_m \in R$.

Using this, we can construct a polynomial $g \in I'$ that has the same degree and same leading coefficient as f_{m+1}, namely,

$$g := \sum_{i=1}^{m} h_i f_i x^{\deg f_{m+1} - \deg f_i}$$

Then $f_{m+1} - g$ has degree strictly smaller than f_{m+1} and lies in $I \setminus I'$ (why?). But this contradicts our choice of f_{m+1} as a smallest-degree element. □

Exercise 5.6. *For every n, find an ideal in $\mathbb{R}[x, y]$ that needs at least n generators.*

6 The Nullstellensatz

The German word *Nullstellensatz*, meaning "zero locus theorem," is commonly used in English to denote a basic and classical result of algebraic geometry. It applies to varieties over an algebraically closed field, most notably over $\mathbb{C}$—a very important assumption.

[3] Here "basis" refers to what we call "generating set." In linear algebra, bases are inclusion-minimal generating sets and they have a number of neat properties, such as all having the same size for a given vector space. In contrast, different inclusion-minimal generating sets of an ideal may have very different sizes and thus, for example, they are unsuitable for defining "dimension."

For a field $\mathbb{k}$ that is not algebraically closed, one can sometimes obtain useful information by applying the Nullstellensatz with the **algebraic closure** $\overline{\mathbb{k}}$ of $\mathbb{k}$, which is an inclusion-minimal algebraically closed field extending $\mathbb{k}$; as it turns out, $\overline{\mathbb{k}}$ is determined uniquely up to isomorphism.

Exercise 6.1. (a) *Prove that for every field $\mathbb{k}$, possibly finite, there are infinitely many irreducible polynomials in $\mathbb{k}[x]$, none a constant multiple of another. (Recall that a polynomial f is* **irreducible** *if it is not a product $f = gh$ with $\deg g, \deg h \geq 1$.)*

(b) *Deduce that every algebraically closed field is infinite.*

The weak Nullstellensatz: a theorem about alternative. There are several ways of stating the Nullstellensatz. The following one is perhaps the most intuitive. It is called "weak" but the full version can be derived from it fairly quickly.

Many areas of mathematics have *theorems of alternative*, with the following structure: *if something cannot be done, then this impossibility must be caused by an "obvious" obstacle.* In linear algebra, for example, if a system $Ax = b$ of linear equations has no solution, then there is a linear combination of the equations that has all coefficients on the left-hand side zero and the right-hand side nonzero. In other words, there exists a vector y such that $y^T A = 0$ and $y^T b \neq 0$.

Exercise 6.2. *Prove this using suitable theorems from linear algebra.*

The weak Nullstellensatz can also be stated in this form: if a system of polynomial equations $f_1 = f_2 = \cdots = f_m = 0$, with $f_1, \ldots, f_m \in \mathbb{k}[x_1, \ldots, x_n]$ and $\mathbb{k}$ *algebraically closed*, has no solution, then there are polynomials $h_1, \ldots, h_m \in \mathbb{k}[x_1, \ldots, x_n]$ such that $h_1 f_1 + \cdots + h_m f_m = 1$. The last equation is an obvious reason of unsolvability of the original system, since any common zero of $f_1, \ldots, f_m$ would also be a zero of $h_1 f_1 + \cdots + h_m f_m$, but the latter is never zero.

Here is the usual, formally somewhat different, statement.

Theorem 6.3 (Weak Nullstellensatz)**.** *Let $\mathbb{k}$ be algebraically closed and let I be an ideal in $\mathbb{k}[x_1, \ldots, x_n]$ such that $V(I) = \emptyset$; that is, there is no common zero. Then $I = \langle 1 \rangle = \mathbb{k}[x_1, \ldots, x_n]$.*

Exercise 6.4. (a) *Give an example of how this fails over $\mathbb{R}$.*

(b) *Prove the weak Nullstellensatz for $n = 1$.*

The usual proofs of the weak Nullstellensatz, including those given here, are nonconstructive—they do not provide the h_i for given f_i. Algorithmic methods exist as well, and we will mention them later on. But it should be said that although the weak Nullstellensatz provides an "obvious" reason, or proof, of unsolvability of a given polynomial system, that proof is not necessarily very compact. Indeed, examples are known in which the smallest possible degree of the h_i has to be exponential in n (see [Kol88] for precise bounds).

The ideal–variety correspondence: the (strong) Nullstellensatz. The strong Nullstellensatz basically says that, over an algebraically closed field, algebraic varieties in $\Bbbk^n$ are in one-to-one correspondence with ideals in $\Bbbk[x_1, \ldots, x_n]$. Or actually, not with all ideals but **radical** ones, where an ideal I is radical if $f^s \in I$ for some natural number s implies $f \in I$.

This extra condition is needed since, e.g., the ideals $\langle x \rangle$ and $\langle x^2 \rangle$ in $\mathbb{C}[x]$ both define the same variety, namely $\{0\}$—but only the first one is radical. For an arbitrary ideal I in a ring R, its **radical** $\sqrt{I}$ is defined in the expected way, as $\{f \in R : f^s \in I \text{ for some } s\}$.

Exercise 6.5. *Check that $\sqrt{I}$ is an ideal.*

For a set $S \subseteq \Bbbk^n$, let

$$I(S) := \{f \in \Bbbk[x_1, \ldots, x_n] : f \text{ vanishes on } S\};$$

clearly, this is an ideal.

Exercise 6.6. (a) *Check that $V(I(X)) = X$ for every variety X, over any field.* (b) *Verify that $\sqrt{I} \subseteq I(V(I))$ for every ideal $I \subseteq \Bbbk[x_1, \ldots, x_n]$, again over any field.*

Theorem 6.7 (Strong Nullstellensatz)**.** *Let $\Bbbk$ be algebraically closed and let I be an ideal in $\Bbbk[x_1, \ldots, x_n]$. Then $I(V(I)) = \sqrt{I}$. Thus, if $f_1, \ldots, f_m \in \Bbbk[x_1, \ldots, x_n]$ are polynomials and g is a polynomial that vanishes on $V(f_1, \ldots, f_m)$, then then there are an integer s and polynomials $h_1, \ldots, h_m \in \Bbbk[x_1, \ldots, x_n]$ such that $g^s = \sum_{i=1}^m h_i f_i$.*

Proof of the strong Nullstellensatz from the weak one. This is known as the *Rabinowitsch trick*: we add a new variable and a new equation to get an unsatisfiable system, for which we apply the weak Nullstellensatz.

Namely, let $I = \langle f_1, \ldots, f_m \rangle$; then the polynomials $f_1, \ldots, f_m$ and $x_{n+1}g - 1 \in \Bbbk[x_1, \ldots, x_{n+1}]$ have no common zero in $\Bbbk^{n+1}$. So by the weak Nullstellensatz we have

$$h_1 f_1 + \cdots + h_m f_m + h_{m+1}(x_{n+1}g - 1) = 1 \tag{1}$$

for some $h_1, \ldots, h_{m+1} \in \Bbbk[x_1, \ldots, x_{n+1}]$.

This equality holds for every value of the variables, and in particular, with $x_{n+1} = 1/g(x_1, \ldots, x_n)$ whenever $g(x_1, \ldots, x_n) \neq 0$. Hence the rational function resulting by substituting $x_{n+1} = 1/g(x_1, \ldots, x_n)$ into the left-hand side of (1) equals 1 whenever $g \neq 0$.

We multiply both sides of the resulting equality by g^s, where s is the highest power of x_{n+1} appearing in (1). This yields the following equality of polynomial functions $\Bbbk^n \to \Bbbk$

$$h'_1 f_1 + \cdots + h'_m f_m = g^s \tag{2}$$

which holds at all points except possibly at the zeros of g (here $h'_1, \ldots, h'_m \in \Bbbk[x_1, \ldots, x_n]$; also note that the term $h_{m+1}(x_{n+1}g - 1)$ vanishes). Using the fact that every algebraically closed field is infinite (Exercise 6.1) and, for example, the Schwartz–Zippel theorem, we get that (2) holds as an equality of polynomials, and this concludes the proof. □

The strong Nullstellensatz shows that, with $\Bbbk$ algebraically closed, an algebraic variety in $\Bbbk^n$ and a radical ideal in $\Bbbk[x_1, \ldots, x_n]$ are just two ways of looking at the same object. Such alternative views of mathematical objects are often very useful.

Several proofs of the Nullstellensatz are known, usually with numerous variations. Here we essentially follow a particularly simple proof from [Arr06], in which we meet a classical tool—resultants.

6.1 Intermezzo: Resultants

Resultants are useful for several purposes. They provide a way of detecting when two polynomials have a nonconstant common factor (or, over an algebraically closed field, a common root), and they are useful for eliminating variables from a polynomial system or, in geometric terms, for projecting an algebraic variety on a coordinate subspace. Here we introduce them briefly, aiming mainly at the properties we will really use.

Excluding common zero of two polynomials. For a while we will be dealing with *univariate* polynomials f and g; first let us assume that they are over a field $\Bbbk$. Let us write them as $f(x) = \sum_{i=0}^{k} f_i x^i$ and $g(x) = \sum_{j=0}^{\ell} g_j x^j$.

To see how one can naturally arrive at the resultant, let us consider the system of two polynomial equations $f = 0$, $g = 0$. A possible way of showing that it is unsolvable, i.e., f and g have no common root, is to find polynomials $a, b \in \Bbbk[x]$ such that the polynomial $af + bg$ is a nonzero constant, say 1.

First we observe that if such a and b exist, we may as well assume $\deg a < \ell$ and $\deg b < k$. This is because if some $af + bg = 1$, then also $a'f + b'g = 1$, where $a' = a + pg$ and $b' = b - pf$ for some $p \in \Bbbk[x]$. Hence we can reduce a modulo g to have degree smaller than ℓ, and then b must have degree smaller than k, for otherwise, we would have $\deg bg \geq k+\ell > \deg af$, and so $af+bg = 1$ would be impossible.

Let us regard the coefficients of a and b as above as unknowns. The requirement $af + bg = 1$ is an equality of polynomials of degree at most $k + \ell$. By comparinge the coefficients of each of the relevant powers of x on both sides, we obtain a system of $k + \ell$ linear equations with $k + \ell$ unknowns. The reader may want to write this system down and see that its matrix looks as follows (we show it for the special case $k = 5$ and $\ell = 3$, which makes clear what the general case is):

$$\begin{pmatrix}
f_0 & 0 & 0 & g_0 & 0 & 0 & 0 & 0 \\
f_1 & f_0 & 0 & g_1 & g_0 & 0 & 0 & 0 \\
f_2 & f_1 & f_0 & g_2 & g_1 & g_0 & 0 & 0 \\
f_3 & f_2 & f_1 & g_3 & g_2 & g_1 & g_0 & 0 \\
f_4 & f_3 & f_2 & 0 & g_3 & g_2 & g_1 & g_0 \\
f_5 & f_4 & f_3 & 0 & 0 & g_3 & g_2 & g_1 \\
0 & f_5 & f_4 & 0 & 0 & 0 & g_3 & g_2 \\
0 & 0 & f_5 & 0 & 0 & 0 & 0 & g_3
\end{pmatrix}.$$

This is called the **Sylvester matrix** of f and g.

The **resultant** of f and g, denoted by $\mathrm{Res}(f, g, x)$, is the determinant of the Sylvester matrix, which is an element of $\Bbbk$. From the above discussion it is clear that if $\mathrm{Res}(f, g, x) \neq 0$, then the considered linear system has a solution, and so the desired a and b with $af + bg = 1$ exist, witnessing the nonexistence of a common root of f and g.

Exercise 6.8. (a) *Using Euclid's algorithm, check that if $f, g \in \Bbbk[x]$ have no nonconstant common factor, then there are polynomials $u, v \in \Bbbk[x]$ with $uf + vg = 1$. (The reverse implication is obvious.)*

(b) *Using (a), prove that for $f, g \in \Bbbk[x]$, where $\Bbbk$ need not be algebraically closed, $\mathrm{Res}(f, g, x) = 0$ implies that f and g have a nonconstant common factor.*

Resultant over a ring. We will need a slightly more general setting, where $f, g \in R[x]$ are polynomials over a ring R (commutative with 1 as usual). The definition above still makes sense and $\mathrm{Res}(f, g, x)$ is an element of R. The next lemma, which we will need later, provides another way of showing that if $\mathrm{Res}(f, g, x) \neq 0$, then f and g have no common root.

Lemma 6.9. *For every $f, g \in R[x]$, $\deg f = k$, $\deg g = \ell$, there are $a, b \in R[x]$ with $\deg a \leq \ell - 1$, $\deg b \leq k - 1$, and $\mathrm{Res}(f, g, x) = af + bg$.*

Proof. We do the following row operations on the Sylvester matrix: for $i = 2, 3, \ldots, k + \ell$ we add the ith row multiplied by x^{i-1} to the first row. After that the first row is

$$(f, xf, \ldots, x^{\ell-1}f, g, xg, \ldots, x^{k-1}g).$$

Expanding this determinant, which still equals $\mathrm{Res}(f, g, x)$, according to the first row, we obtain precisely an expression of the desired form $af + bg$ with a, b as in the lemma. □

6.2 Proof of the weak Nullstellensatz

We need a lemma saying that in a polynomial of degree d, we can make the coefficient of x_1^d nonzero by a suitable invertible linear substitution. This result is quite simple; it is a special case of a more intricate result known as the *Noether normalization lemma.*

Lemma 6.10. *If $\Bbbk$ is an infinite field and $f \in \Bbbk[x_1, \ldots, x_n]$ is a polynomial of degree $d \geq 1$, then there are $\lambda_1, \ldots, \lambda_{n-1} \in \Bbbk$ such that the coefficient of x_n^d in the polynomial $f'(x_1, \ldots, x_n) := f(x_1 + \lambda_1 x_n, \ldots, x_{n-1} + \lambda_{n-1} x_n, x_n)$ is nonzero.*

Proof. Let f_d denote the sum of all terms of degree d in f (this is called the *homogeneous component* of f of degree d). Then the coefficient of x_n in f' equals $f_d(\lambda_1, \ldots, \lambda_{n-1}, 1)$. Since $\Bbbk$ is infinite, the nonzero polynomial $f_d(x_1, \ldots, x_{n-1}, 1)$ cannot vanish everywhere on $\Bbbk^{n-1}$. □

Proof of the weak Nullstellensatz. We establish the contraposition, so we assume that I is an ideal properly contained in $\Bbbk[x_1, \ldots, x_n]$, and we want to find a common zero $(a_1, \ldots, a_n) \in \Bbbk^n$ of all polynomials in I.

We proceed by induction on n, considering the case $n = 1$ settled (Exercise 6.4). So let $n > 1$.

By Lemma 6.10 we can make a change of variables so that I contains a polynomial g of degree $d \geq 1$ with the term x_n^d. Since this substitution is invertible, if we find a common zero for the ideal obtained from I after the substitution, we can convert it back to a common zero for the original I. So we assume we have $g \in I$ as above.

Let I' be the set of all polynomials in I that do not contain the variable x_n (that is, there is no term with nonzero coefficient and nonzero power of x_n). We can regard I' as a subset of $\Bbbk[x_1, \dots, x_{n-1}]$; then it is a proper ideal (right?), and so by the inductive hypothesis, there is $(a_1, \dots, a_{n-1})$, a common zero of all polynomials in I'.

Now we claim that the set

$$J := \{f(a_1, \dots, a_{n-1}, x_n) : f \in I\},$$

which is obviously an ideal, is not all of $\Bbbk[x_n]$. Once we prove this claim, we will be done, since by the 1-dimensional weak Nullstellensatz all polynomials in J have a common zero $a \in \Bbbk$, and then $(a_1, \dots, a_{n-1}, a)$ is a common zero for I.

To prove the claim, we need to check that $1 \notin J$, so for contradiction, we assume that there is $f \in I$ such that $f(a_1, \dots, a_{n-1}, x) = 1$ (this is an equality of univariate polynomials). We fix f, as well as g as above, i.e., of degree d and with term x_n^d.

Let us consider f and g as polynomials in x_n: $f = \sum_{i=0}^k f_i x_n^i$, $g = \sum_{j=0}^d g_j x_n^j$, $f_0, \dots, f_k$, $g_0, \dots, g_d \in R := \Bbbk[x_1, \dots, x_{n-1}]$.

By Lemma 6.9, the resultant $\operatorname{Res}(f, g, x_n) \in R$ can be written as $af + bg$ with $a, b \in R[x]$, and hence it belongs to I'. To finish the proof, we will show that $\operatorname{Res}(f, g, x_n)$ is nonzero at $(a_1, \dots, a_{n-1})$, and hence it cannot belong to I'.

The equality $f(a_1, \dots, a_{n-1}, x) = 1$ means that $f_0(a_1, \dots, a_{n-1}) = 1$ and f_1 through f_k vanish at $(a_1, \dots, a_{n-1})$. Also, by the choice of g, we have $g_d = 1$ (identically). Looking at the Sylvester matrix of f and g, again for notational simplicity in the particular case $\deg f = 5$, $\deg g = 3$, i.e.,

$$\begin{pmatrix}
f_0 & 0 & 0 & g_0 & 0 & 0 & 0 & 0 \\
f_1 & f_0 & 0 & g_1 & g_0 & 0 & 0 & 0 \\
f_2 & f_1 & f_0 & g_2 & g_1 & g_0 & 0 & 0 \\
f_3 & f_2 & f_1 & g_3 & g_2 & g_1 & g_0 & 0 \\
f_4 & f_3 & f_2 & 0 & g_3 & g_2 & g_1 & g_0 \\
f_5 & f_4 & f_3 & 0 & 0 & g_3 & g_2 & g_1 \\
0 & f_5 & f_4 & 0 & 0 & 0 & g_3 & g_2 \\
0 & 0 & f_5 & 0 & 0 & 0 & 0 & g_3
\end{pmatrix},$$

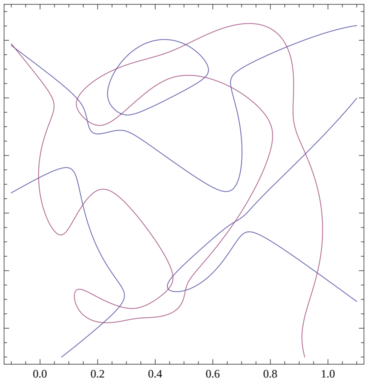
we see that at $(a_1, \dots, a_{n-1})$ it is an upper triangular matrix with 1s on the main diagonal, and hence $\operatorname{Res}(f, g, x_n)(a_1, \dots, a_{n-1}) = 1$. □

7 Bézout's inequality in the plane

One of the questions that often comes up in applications is, given a system of polynomial equations $f_1 = 0, \ldots, f_m = 0$, $f_1, \ldots, f_m \in \Bbbk[x_1, \ldots, x_n]$, what can be said about the existence and number of solutions? In order to avoid trivialities, we always assume that $d_i := \deg f_i \geq 1$ for all i.

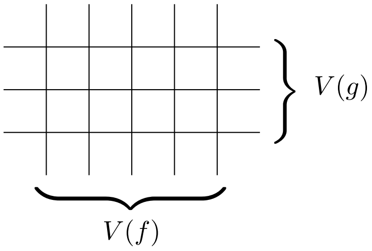
In general this is not an easy question, and in this section we will consider the special case with two equations $f(x, y) = 0$, $g(x, y) = 0$ in two variables, which is considerably simpler than the general setting but still interesting. (We are leaving aside the case of a single equation $f = 0$, which has already been treated to some extent, at least implicitly.)

Here is an example of the zero set of two polynomials f and g of degree 5; each of them has been created by passing the zero set through 25 random points in $[0, 1]^2$ using Lemma 4.2.



The system $f = g = 0$ may have infinitely many solutions—this we can see already in the case of linear equations, where f may be a multiple of g or, speaking geometrically, the two lines described by the equations may coincide. For two polynomials, infinitely many solutions may occur if f and g have a nonconstant common factor, and for $\Bbbk$ algebraically closed they actually do occur in such case.

As we will see, excluding a common factor leads to finitely many solutions. Let us consider an example with $f(x, y) = (x-1)(x-2)\cdots(x-k)$ and $g(x, y) = (y-1)(y-2)\cdots(y-\ell)$. We have $\deg f = k$, $\deg g = \ell$, and the zero sets look like this:

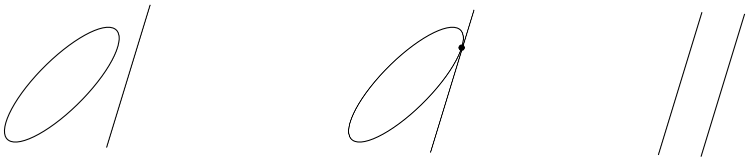


Thus, there can be as many as $k\ell$ distinct solutions. This example, trivial as it may look, is actually quite useful: the union of k hyperplanes is the zero set of a degree-k polynomial, and although this is not really a typical polynomial, it can serve for a quick sanity check of many things.

The following theorem asserts that, assuming no common factor, we cannot have more solutions than in the example.

Theorem 7.1 (Bézout's inequality in the plane)**.** *Let $f, g \in \mathbb{k}[x, y]$ be polynomials of degrees $k, \ell \geq 1$, respectively, having no nonconstant common factor. Then $|V(f, g)| \leq k\ell$.*

In algebraic geometry, Bézout's theorem is often stated as an equality: under suitable assumptions, there are *exactly* $k\ell$ solutions. The assumptions have to address three issues: (a) the field has to be algebraically closed; (b) we have to count solutions with appropriately defined multiplicity; and (c) we also have to count solutions "at infinity." The next drawing illustrates these issues:



not algebraically closed multiplicity at infinity

We will talk about solutions at infinity later. Handling multiplicities properly takes a substantial amount of work, and we will not consider it here. However, Bézout's theorem is usually applied in the inequality form.

Theorem 7.1 can be proved in several ways, for example using resultants. The proof shown below is ingenious, short, and introduces a general approach used for handling the concept of dimension in algebraic geometry. We begin with some general considerations.

Coordinate rings, and measuring them. We recall from algebra that if I is an ideal in a (commutative) ring S, we can form the *quotient ring* S/I, whose elements are equivalence classes of elements of S, with a, b equivalent if $a - b \in I$. Here we will consider the case where S is the polynomial ring $\mathbb{k}[x_1, \ldots, x_n]$ and I is an ideal in S.

In particular, if $X \subseteq \mathbb{k}^n$ is an algebraic variety and $I = I(X) \subseteq \mathbb{k}[x_1, \ldots, x_n]$, then the quotient ring $\mathbb{k}[x_1, \ldots, x_n]/I$ is called the **coordinate ring** of X and denoted by $\mathbb{k}[X]$. It has an intuitive meaning: its elements can be represented by polynomials, but two polynomials are considered the same if they coincide on X (strictly speaking, this is literally true only over infinite fields).

Being determined by the ideal $I = I(X)$, the coordinate ring carries the same information as I, but some things are more convenient to express in terms of the coordinate ring. Moreover, $\mathbb{k}[X]$ is more suitable for representing the variety X "up to isomorphism" (which we are going to define later).

Now we want to measure the "size" of coordinate rings. Slightly more generally, we consider an ideal I and the quotient ring $R := \mathbb{k}[x_1, \ldots, x_n]/I$. They are both closed under addition and under multiplication by elements of $\mathbb{k}$, and so they are also vector spaces over $\mathbb{k}$.

The vector space dimension of R, or of I, in itself is usually not a very good measure of "size," since it is most often infinite. Certainly, for $I = I(X)$ and $R = \Bbbk[X]$, it does not capture the intuitive geometric notion of dimension of the variety X. The trick is to consider subspaces consisting of polynomials up to some given degree d.

For the ideal I this can be done in the obvious manner: we let $I_{\le d}$ consist of all polynomials in I of degree at most d. For R this is slightly more tricky, since two polynomials representing the same element of R may have different degrees.

We thus define R_d as the quotient vector space $\Bbbk[x_1, \dots, x_n]_{\le d}/I_{\le d}$, so the elements of R_d are represented by polynomials of degree at most d, with the same equivalence as that for R.

By a well known fact from linear algebra about quotient spaces, we have $\dim R_d + \dim I_{\le d} = \dim \Bbbk[x_1, \dots, x_n]_{\le d} = \binom{n+d}{n}$, the last equality being Fact 4.1. In particular, R_d and $I_{\le d}$ have finite dimension for every d.

The vector-space dimension of $\Bbbk[X]_d$, considered as a function of d, carries a lot of information about the variety X, and it has a name—again after Hilbert.

Let $R = \Bbbk[x_1, \dots, x_n]/I$ be a quotient of the polynomial ring $\Bbbk[x_1, \dots, x_n]$, and let R_d be the vector space defined as above. Then the **Hilbert function** of R (or, for $I = I(X)$, also of X) is defined as

$$\mathrm{HF}_R(d) := \dim R_d.$$

If $I \subseteq I' \subseteq \Bbbk[x_1, \dots, x_n]$ are ideals and R, R' are the corresponding quotient rings, we have $\mathrm{HF}_R \ge \mathrm{HF}_{R'}$ (this follows from $\mathrm{HF}_R(d) = \binom{n+d}{n} - \dim I_{\le d}$). For varieties this yields $\mathrm{HF}_X \le \mathrm{HF}_{X'}$ for $X \subseteq X'$, which we will freely use in the sequel.

Proof of Theorem 7.1. The plan for proving the planar Bézout inequality is now this:

(i) We check that if $X \subseteq \Bbbk^2$ is an m-point set, then the Hilbert function of X is at least m for all sufficiently large d.

(ii) We show that if f and g have no nonconstant common factor, then $\mathrm{HF}_R(d) \le k\ell$, where $R := \Bbbk[x, y]/\langle f, g\rangle$, again for sufficiently large d.

To prove (i), let $X = \{a_1, \dots, a_m\} \subset \Bbbk^2$, and let us choose a system $\varphi_1, \dots, \varphi_m$ of functions $X \to \Bbbk$ that are linearly independent. For example, we can set $\varphi_i(a_j) := \delta_{ij}$, the Kronecker delta, with $\delta_{ii} = 1$ and $\delta_{ij} = 0$ for $i \ne j$.

According to Exercise 4.4(a), for each φ_i there is a polynomial $p_i \in \Bbbk[x, y]$ whose values on X coincide with φ_i. Then the p_i are linearly independent as elements of the coordinate ring $\Bbbk[X]$, and this proves $\dim \Bbbk[X]_d \ge m$ for all $d \ge \max \deg p_i$. (This argument works for any number of variables, not only two.)

As for (ii), let us first consider the ideals $K := \langle f\rangle$ and $L := \langle g\rangle$. We claim that for $d \ge k$, we have $\dim K_{\le d} = \dim \Bbbk[x, y]_{\le d-k}$. This is because every polynomial in $p \in K$ has the form $p = af$, and p determines a uniquely. (Here we use that $\Bbbk[x,y]$ is a *unique factorization domain*; Exercise 7.3 below.) Of course, we also have $\dim L_{\le d} = \dim \Bbbk[x,y]_{d-\ell}$ for $d \ge \ell$.

What we want to bound is $\dim I_{\le d}$, where $I = \langle f, g\rangle$. We have $I = \{af+bg : a, b \in \Bbbk[x,y]\} = \{p+q : p \in K, q \in L\}$. The sum of two polynomials of degree at most d again has degree at most d, and hence $I_{\le d} \supseteq K_{\le d} + L_{\le d}$.

Exercise 7.2. *Find an example where this inclusion is proper.*

Fortunately, since we need to bound $\dim R_d$ from above and thus $\dim I_{\le d}$ from below, the inclusion goes in the right direction. By the well-known formula for the dimension of a sum of vector spaces, we have

$$\dim(K_{\le d} + L_{\le d}) = \dim K_{\le d} + \dim L_{\le d} - \dim(K \cap L)_{\le d}.$$

It remains to note that, since f and g have no common factor, a polynomial divisible by both f and g must be divisible by fg, and so $K \cap L = \langle fg \rangle$. Hence $\dim(K \cap L)_{\le d} = \dim \Bbbk[x,y]_{\le d-k-\ell}$ for $d \ge k+\ell$.

The rest is calculation with binomial coefficients:

$$\begin{aligned}\dim R_d &= \dim \Bbbk[x,y]_{\le d} - \dim I_{\le d}\\ &\le \dim \Bbbk[x,y]_{\le d} - \dim(K_{\le d} + L_{\le d})\\ &= \binom{d+2}{2} - \binom{d-k+2}{2} - \binom{d-\ell+2}{2} + \binom{d-k-\ell+2}{2}\\ &= k\ell\end{aligned}$$

(assuming $d \ge k+\ell$). □

Exercise 7.3. *We recall that a (commutative) ring R is called an* **integral domain** *if the product of every two nonzero elements is nonzero. An element $a \in R$ is* irreducible *if it cannot be written as a product $a = bc$ with neither b nor c invertible.*

(a) *Let R be an integral domain in which every nonzero element has a unique factorization into irreducibles (unique up to reordering and multiplication by invertible elements). The* contents $\mathrm{cont}(f)$ *of a polynomial $f \in R[x]$ is defined as the greatest common divisor of all coefficients of f. Show that* $\mathrm{cont}(fg) = \mathrm{cont}(f)\,\mathrm{cont}(g)$.

(b) *Prove that every univariate polynomial $f \in \Bbbk[x]$ over a field has a unique factorization into irreducible polynomials.*

(c) *Prove by induction on n that every $f \in \Bbbk[x_1, \ldots, x_n]$ has a unique factorization into irreducible polynomials.*

8 More properties of varieties

In this section we introduce further basic notions and results concerning algebraic varieties. Building this theory properly with all details requires much more space, and so we try to present a reasonable selection. We will encounter many clever and sophisticated notions, and one should not expect to master all of them quickly, but hopefully they will look less frightening next time. Reading this section should give some first impression and basic vocabulary; for serious work one should study a proper textbook.

8.1 Irreducible components

Irreducible varieties. The union of the x-axis and y-axis in the plane is an algebraic variety, namely, $V(xy)$, which can naturally be decomposed into two proper subvarieties, $V(x)$ and $V(y)$. Varieties that cannot be further decomposed are called irreducible:

A variety $X \subseteq \Bbbk^n$ is **irreducible** if we cannot express $X = X_1 \cup X_2$ with X_1 and X_2 both varieties and proper subsets of X.

As we have remarked, some sources even reserve the term variety only for irreducible varieties, and irreducibility is extremely important. We have already seen a hint of this in Bézout's inequality, and many other theorems require irreducibility assumptions. For example, it turns out that an irreducible variety over an algebraically closed field has the same "local dimension" in the neighborhood of each point (we have not yet defined dimension rigorously, but surely the reader has some intuitive idea), while a reducible variety may be, e.g., the union of a plane and a line.

Exercise 8.1. (a) *(Any field) Show that if a variety $X \subseteq \Bbbk^n$ is irreducible, then $I = I(X)$ is a* **prime** *ideal; that is, $fg \in I$ implies $f \in I$ or $g \in I$.*

(b) *(Algebraically closed field) Prove that if $X \subseteq \Bbbk^n$ is a variety with $\Bbbk$ alrebraically closed such that $I(X)$ is prime, then X is irreducible.*

(c) *Check that a prime ideal is radical, but not necessarily the other way around.*

Proposition 8.2. *Every variety X can be decomposed as a finite union $X = X_1 \cup \cdots \cup X_k$ of irreducible varieties. Moreover, assuming that $X_i \not\subseteq X_j$ for all $i \neq j$, the decomposition is unique up to reordering.*

The X_i as in the proposition are called the **irreducible components** of X.

Sketch of proof. Finiteness follows from the Hilbert basis theorem: if we could keep decomposing indefinitely, we would obtain an infinite descending chain of varieties $X_1 \supsetneq X_2 \supsetneq X_3 \supsetneq \cdots$, whose corresponding ideals would form an infinite ascending chain, and this is impossible since $\Bbbk[x_1, \ldots, x_n]$ is Noetherian.

As for uniqueness, assuming two minimal decompositions into irreducibles $X = X_1 \cup \cdots \cup X_k = X'_1 \cup \cdots \cup X'_\ell$, we observe that if some X_i were not among the X'_j, then the $X_i \cap X'_j$ would properly decompose X_i or vice versa. □

One of the basic sources of difficulties in algebraic geometry is that *the intersection of irreducible varieties need not be irreducible.* A simple example is with two irreducible algebraic curves in $\Bbbk^2$ intersecting in at least two points, but there are more interesting higher-dimensional examples as well, one of them to be mentioned in Section 9 below.

We also stress that the task of finding the irreducible decomposition of a given variety is highly nontrivial in general, although algorithmically solvable.

The Zariski topology. In the language of algebraic geometry, a set $S \subseteq \Bbbk^n$ is called **Zariski closed** or just **closed** if it is a variety, and it is **(Zariski) open** if its complement is a variety. Readers familiar with the notion of topological space can check that this defines a topology on $\Bbbk^n$, although a somewhat peculiar one. Nonempty open sets are very big (assuming an infinite field), they are dense in $\Bbbk^n$ and every two intersect. Thus, the topology is not Hausdorff. Yet it provides a convenient framework and terminology.

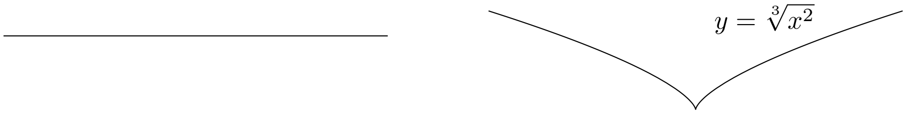
Exercise 8.3. *Let $X \subseteq \Bbbk^m$ and $Y \subseteq \Bbbk^n$ be irreducible varieties. Prove that the product $X \times Y \subseteq \Bbbk^{m+n}$ is irreducible as well.*

8.2 Morphisms of affine varieties

Having defined a class of objects, affine algebraic varieties in our case, one should ask what is an appropriate notion of *morphisms* of the objects. Familiar examples of morphisms include linear maps of vector spaces, homomorphisms of groups, rings, fields, but also of graphs, and continuous maps of topological spaces.

For affine algebraic varieties, morphisms are called *regular maps.* A *polynomial map* $f\colon \Bbbk^m \to \Bbbk^n$ is a map $f = (f_1, \ldots, f_n)$ such that each f_i is given by a polynomial in $\Bbbk[x_1, \ldots, x_m]$. If $X \subseteq \Bbbk^m$ and $Y \subseteq \Bbbk^n$ are varieties, then a *regular map* $f\colon X \to Y$ is a map that is a restriction of a polynomial map $\bar{f}\colon \Bbbk^m \to \Bbbk^n$ to X and satisfies $f(X) \subseteq Y$.

An **isomorphism** of affine varieties is a regular map with a regular inverse. While the affine line $\mathbb{R}$ is homeomorphic as a topological space to the "cusp curve" $V(x^2 - y^3)$, it can be shown that they are not isomorphic as affine varieties.



We note that if $f\colon X \to Y$ is a regular map and $\varphi\colon Y \to \Bbbk$ is a polynomial function on Y, i.e., an element of the coordinate ring $\Bbbk[Y]$, then the composition $\varphi f\colon X \to \Bbbk$ belongs to $\Bbbk[X]$. Thus, the composition with f induces a mapping $f^*\colon \Bbbk[Y] \to \Bbbk[X]$ (note the change of direction compared to f!). Moreover, f^* is a $\Bbbk$-algebra homomorphism, meaning that it is a ring homomorphism for which, in addition, $f^*(\alpha) = \alpha$ for every $\alpha \in \Bbbk$.

Conversely, it is not hard to show that every $\Bbbk$-algebra homomorphism $\Bbbk[Y] \to \Bbbk[X]$ equals f^* for some regular map $f\colon X \to Y$.

Exercise 8.4. *Prove that; start with $X = \Bbbk^m$, $Y = \Bbbk^n$.*

It follows that two varieties are isomorphic exactly if their coordinate rings are isomorphic as $\Bbbk$-algebras. So the coordinate ring provides a "coordinate-free" representation of a variety, independent of a specific embedding of the variety in some $\Bbbk^n$.

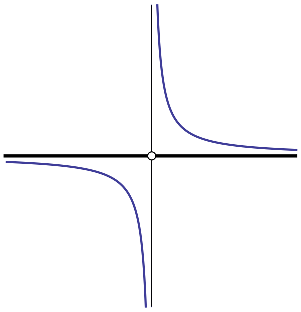
A useful way of proving irreducibility. Let $X \subseteq \Bbbk^m$ be an irreducible variety, and let $f\colon X \to \Bbbk^n$ be a regular map. Then it is easy to check that the image $f(X)$ is irreducible, but the statement has to be understood in a right way.

Indeed, as we will discuss below in more detail, $f(X)$ need not be a variety! So we generalize irreducibility to an arbitrary set $S \subseteq \Bbbk^n$, meaning that we cannot write $S = (S \cap X_1) \cup (S \cap X_2)$, where X_1, X_2 are varieties and $S \cap X_1 \neq S \neq S \cap X_2$. Then we can see that if $f(X)$ were reducible, then so would be X, because the preimage of a variety under a regular map is always a variety, as is easy to check.

Thus, in particular, if we can express some variety Y parametrically, as the image of some $\Bbbk^m$, or of some other irreducible variety X, under a polynomial map f, then Y is irreducible. More generally, it suffices that the image $f(X)$ be **Zariski dense** in Y, meaning that Y is the smallest variety containing $f(X)$.

As an example, let m, n and $r \leq \min(m,n)$ be natural numbers, and consider the **determinantal variety** $D_r(m,n)$ consisting of all $m \times n$ matrices, considered as points in $\Bbbk^{mn}$, that have rank strictly smaller than r. This is indeed a variety since the rank condition can be expressed as vanishing of all $r \times r$ minors. Since an $m \times n$ matrix A has rank at most $r-1$ iff it can be expressed as a product UV, where U is $m \times (r-1)$ and V is $(r-1) \times n$, we have a surjective regular map $\Bbbk^{(r-1)(m+n)} \to D_r(m,n)$, and hence the determinantal variety is irreducible.

Projections and images of affine varieties: constructible sets. Let us consider the variety $X := V(xy-1)$, a hyperbola, and project it onto the x-axis:



The projection $\pi(X)$ is the x-axis minus 0, certainly not an algebraic variety. Passing to an algebraically closed setting, complex numbers, does not help—the 0 is still missing. So affine algebraic varieties are not closed under projections, and under regular maps in general.

One remedy is to add points at infinity and work in the projective space—see Section 8.5 below. Another approach is to consider a larger class consisting of all sets obtainable from varieties by finitely many set-theoretical operations; these are called **constructible sets**. Using the fact that varieties are closed under intersections and finite unions, it is not difficult to check that every constructible set can be written as

$$(X_1 \setminus Y_1) \cup \cdots \cup (X_k \setminus Y_k),$$

for varieties $X_1, Y_1, \dots, X_k, Y_k$, where we may assume the X_i irreducible and $Y_i \subsetneq X_i$. Then Y_i can be regarded as a set of "exceptional points" in X_i; as we will discuss in Section 8.3, it has smaller dimension than X_i.

We state the following result without proof:

Theorem 8.5 (Chevalley's theorem)**.** *Let $\Bbbk$ be an algebraically closed field, and let $\pi\colon \Bbbk^{m+n} \to \Bbbk^n$ denote the projection on the last n coordinates. Then $\pi(Z)$ is a constructible set for every constructible set $Z \subseteq \Bbbk^{m+n}$ and, in particular, for every variety Z.*

This is actually a result about *quantifier elimination* in the theory of algebraically closed fields, and a nice proof can be found in [MO02].

Corollary 8.6. *The image of a constructible set $Z \subseteq \Bbbk^m$ under a regular map $f\colon \Bbbk^m \to \Bbbk^n$ is a constructible set.*

Sketch of proof. This is a generally useful trick: one needs to check that the graph $G := \{(x, f(x)) \in \Bbbk^m \times \Bbbk^n : x \in Z\}$ is a constructible set; then $f(Z) = \pi(G)$ is constructible by Chevalley's theorem. □

Rational maps. A **rational map** $\varphi\colon \Bbbk^m \to \Bbbk^n$ is given by an n-tuple of rational functions

$$\varphi = \Big(\frac{f_1}{g_1}, \dots, \frac{f_n}{g_n}\Big),$$

where $f_1, g_1, \dots, f_n, g_n \in \Bbbk[x_1, \dots, x_m]$ are polynomials, none of the g_i identically zero.

There is a catch: a rational map is not really a map in the usual sense, because it is undefined on the zero sets of the g_i (for this reason, instead the usual mapping arrow $\to$, one uses $\dashrightarrow$ for a rational map). Nevertheless, it is defined on a Zariski open subset of $\Bbbk^m$, and it is still useful.

A rational map $\varphi\colon X \dashrightarrow Y$ of varieties, with $X \subseteq \Bbbk^m$ irreducible and $Y \subseteq \Bbbk^n$ is, similar to regular maps, a restriction of a rational map $\overline{\varphi}\colon \Bbbk^m \dashrightarrow \Bbbk^n$ to X such that $\varphi(X) \subseteq Y$, but with the extra condition that none of the denominators g_i (assuming f_i and g_i having no common factors) vanishes identically on X.

Two rational maps $X \dashrightarrow Y$ are considered equivalent if they agree on a nonempty Zariski open subset of X (they may be defined on different Zariski open subsets of X, though).

We have seen that an algebraic counterpart of regular maps $X \to Y$ are $\Bbbk$-algebra homomorphisms $k[Y] \to \Bbbk[Y]$ of the coordinate rings. Similarly, rational maps $\varphi\colon X \dashrightarrow Y$ of *irreducible* varieties correspond to $\Bbbk$-algebra homomorphisms $\Bbbk(Y) \to \Bbbk(X)$, where $\Bbbk(X)$ is the quotient field of the coordinate ring $\Bbbk[X]$ (which is an integral domain for X irreducible, so a quotient field makes sense).

The corresponding notion of isomorphism is called *birational equivalence*, and it is more permissive than the isomorphism defined by regular maps. For example, it is known, and not extremely difficult to prove, that every variety (over an algebraically closed field) is birationally equivalent to a *hypersurface*, i.e., a variety defined by a single polynomial.

8.3 Dimension and degree

The dimension of algebraic varieties is defined algebraically, and it has several rather different-looking but equivalent definitions. Here we will mention only some of them, and we will not prove their equivalence.

In this section we will assume an algebraically closed field unless stated otherwise. Things are considerably subtler over an arbitrary field, and it is often preferable to work with schemes there, rather than varieties.

Dimension. Here is a definition which is very simple to state, but rather difficult to work with. The **dimension** of a variety X is the largest n such that there is a chain of properly increasing *irreducible* varieties $\emptyset \subsetneq X_0 \subsetneq \cdots \subsetneq X_n \subseteq X$. (In particular, the empty variety $\emptyset$ has dimension -1.)

The idea is that a proper subvariety of an irreducible variety must be of lower dimension; note that the same definition works for finite-dimensional vector spaces. Since, in the algebraically closed case, irreducible varieties correspond to prime ideals (Exercise 8.1), the dimension is also the length of the longest chain of properly nested prime ideals in $I(X)$ (this notion is called the *Krull dimension* of the coordinate ring of X).

With this definition, even $\dim \mathbb{k}^n = n$ is not obvious (but it is true).

A geometric view, and degree. Another, more geometric way is to define the dimension of a variety $X \subseteq \mathbb{k}^n$ as the largest dimension k of a linear subspace $H \subset \mathbb{k}^n$ such that there is a projection $\pi\colon \mathbb{k}^n \to H$ with $\pi(X)$ Zariski dense in H. Here a *projection* is a linear map $\pi\colon \mathbb{k}^n \to \mathbb{k}^n$ such that $\pi \circ \pi = \pi$, and $H = \pi(\mathbb{k}^n)$.

Another, but equivalent, geometric definition of the dimension considers only the the usual projections on all k-dimensional coordinate subspaces.

It turns out that the property of $\pi(X)$ being Zariski dense in $H = \pi(\mathbb{k}^n)$ is generic, in the sense that the set of the π not having this property is negligible: if we parameterize all projections π onto k-dimensional subspaces by suitable coordinates, then those with $\pi(X)$ not Zariski dense in H satisfy a nontrivial polynomial equation.

This point of view also brings us to the notion of degree. For a projection π and a point $y \in H = \pi(\mathbb{k}^n)$, let us consider the number of preimages $|\{x \in X : \pi(x) = y\}|$. It can be shown that for π and y generic, this number is finite and depends only on X. It is called the **degree** of X and denoted by $\deg X$.

There is also a "dual" view: if X is a k-dimensional variety in $\mathbb{k}^n$, then a generic $(n-k-1)$-dimensional affine subspace of $\mathbb{k}^n$ avoids X, while a generic $(n-k)$-dimensional affine subspace intersects it in $\deg X$ points.

Dimension and regular maps. Regular maps do not increase dimension: if X and Y are varieties (over an algebraically closed field) and $f(X) = Y$, or more generally, if $f(X)$ is Zariski dense in Y for a regular map f, then $\dim Y \le \dim X$. Moreover, if we have $\dim f^{-1}(y) = m$ for all y from a Zariski dense subset of Y, then $\dim Y = \dim X - m$. Proofs can be found in many introductory textbooks.

Generalized Bézout. If X, Y are varieties (over an algebraically closed field), then $\deg(X \cap Y) \le (\deg X)(\deg Y)$, which can be seen as a generalization of

Bézout's inequality (see Heintz [Hei83]).

The Hilbert function and the Hilbert polynomial. We recall that the Hilbert function of a variety X is defined as the Hilbert function $\mathrm{HF}_{\Bbbk[X]}$ of its coordinate ring, and the value $\mathrm{HF}_{\Bbbk[X]}(d)$ is the dimension of the vector space $\Bbbk[X]_d$, which consists of polynomials of degree at most d modulo the polynomials in $I(X)$ of degree at most d.

It turns out that for all sufficiently large d, the Hilbert function coincides with a polynomial, called the **Hilbert polynomial** of X. More precisely, for every quotient ring $R = \Bbbk[x_1, \ldots, x_n]/I$ there exist d_0 and a polynomial, denoted by HP_R and obviously uniquely determined, such that $\mathrm{HP}_R(d) = \mathrm{HF}_R(d)$ for all $d \geq d_0$.

This fact, mysterious as it may look, is not difficult. A short algebraic proof can be found, e.g., in [Sch03, Lemma 2.3.3], and below we will provide a geometric picture explaining the polynomial behavior.

The Hilbert polynomial provides a seemingly very different definition of dimension and degree:

The dimension k of an affine algebraic variety X is the degree of its Hilbert polynomial $\mathrm{HP}_{\Bbbk[X]}$, and the degree of X is $k!$ times the leading coefficient of the Hilbert polynomial.

Monomial orderings. For presenting the promised geometric view of the Hilbert function, we first need to define a linear ordering of the monomials in $\Bbbk[x_1, \ldots, x_n]$; this will also be indispensable later, when we briefly discuss computational aspects of ideals and varieties.

One particular ordering which works for our purposes is the **graded lexicographic ordering**: for two monomials $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and x^β, we first compare the degrees, i.e., $\|\alpha\|_1 = \sum_{i=1}^n \alpha_i$ and $\|\beta\|_1$, and if they are equal, we compare the nonnegative integer vectors α and β lexicographically.

More generally, a **monomial ordering** is a linear ordering $\leq$ on $\mathbb{Z}^n_{\geq 0}$ (we identify monomials with their exponent vectors) that is a well-ordering,[4] and such that $\alpha < \beta$ implies $\alpha + \gamma < \beta + \gamma$ for every $\gamma \in \mathbb{Z}^n_{\geq 0}$. For the considerations in this section, we also need the monomial ordering to be **graded**, meaning that $\|\alpha\|_1 < \|\beta\|_1$ implies $\alpha < \beta$.

So we fix a graded monomial ordering $\leq$. Then every polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ has a uniquely determined **leading monomial** $\mathrm{LM}(f)$, the one that is the largest according to the monomial ordering.

For an ideal I in $\Bbbk[x_1, \ldots, x_n]$, we let $\mathrm{LM}(I) := \langle \mathrm{LM}(f) : f \in I \rangle$; this is a *monomial ideal*, meaning that it is generated by monomials (but of course, being an ideal, it also contains polynomials that are not monomials). We should also warn that if I is generated by some polynomials $f_1, \ldots, f_m$, $\mathrm{LM}(I)$ may be larger than $\langle \mathrm{LM}(f_1), \ldots, \mathrm{LM}(f_m) \rangle$—the reader may want to find an example.

The next claim, which we leave as an exercise, shows that, as far as the Hilbert function is concerned, it is enough to deal with monomial ideals.

[4] That is, every nonempty subset has a minimum element.

Exercise 8.7. *Let us fix a graded monomial ordering, let I be an ideal in $\Bbbk[x_1, \ldots, x_n]$, let $I' := \mathrm{LM}(I)$, and let $R := \Bbbk[x_1, \ldots, x_n]/I$ and $R' := \Bbbk[x_1, \ldots, x_n]/I'$ be the corresponding quotient rings.*

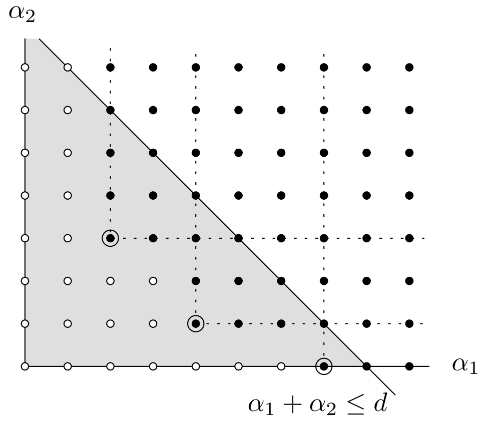
(a) *Show that $I_{\leq d}$ has a basis $(f_1, \ldots, f_m)$ such that $\mathrm{LM}(f_1) > \cdots > \mathrm{LM}(f_m)$, and derive $\dim I_{\leq d} \leq \dim I'_{\leq d}$.*

(b) *Prove that if the f_i constitute a basis of $I_{\leq d}$ as in (a), then $\mathrm{LM}(f_1), \ldots, \mathrm{LM}(f_m)$ generate $I'_{\leq d}$. Conclude that $\mathrm{HF}_{R'} = \mathrm{HF}_R$.*

(c) *Where does the argument use the assumption that the monomial ordering is graded?*

The proof in the exercise also shows that all monomials in $I' = \mathrm{LM}(I)$ are linearly independent, and that each $I'_{\leq d}$ has a basis consisting of monomials.

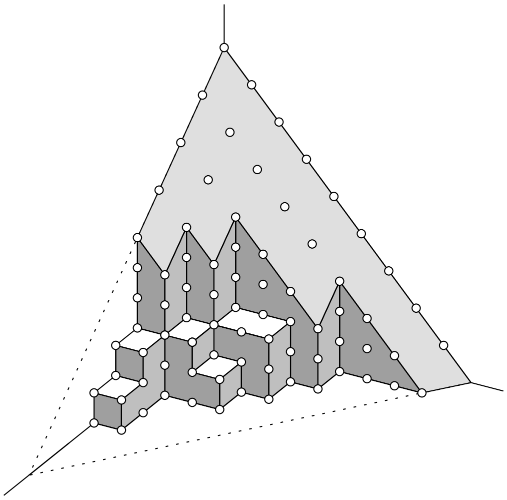
Let us consider $\mathbb{Z}^n_{\geq 0}$, all n-tuples of nonnegative integers, and let us color the exponent vector of every monomial in the monomial ideal I' black; here is a picture for $n = 2$:



Since I' is an ideal, the black dots are the union of finitely many "corners", i.e., translations of the nonnegative orthant—one corner for each generator. The generators are marked by double circles.

The number of black dots in the halfspace $\alpha_1 + \cdots + \alpha_n \leq d$ is the vector-space dimension of $I'_{\leq d}$ (since the corresponding monomials form a basis), and hence the value of $\mathrm{HF}_{R'}(d)$ is the number of white dots in that halfspace (because $\mathrm{HF}_{R'}(d) = \binom{n+d}{d} - \dim I'_{\leq d}$; we do *not* claim that the corresponding monomials form a basis).

From this interpretation one can see why the Hilbert function eventually becomes a polynomial: the key observation is that if we ignore a finite number of "irregular" white dots near the origin, the remaining white dots can be organized into finitely many disjoint axes-parallel "orthants" of various dimensions (semiinfinite rays, quadrants of planes, octants of 3-dimensional subspaces, etc.); this is not quite a proof but almost. The following 3-dimensional picture illustrates how the halfspace $\sum \alpha_i \leq d$ sweeps the set of white dots, after it has already passed the irregular part:



Finally, let us see why the growth of the Hilbert polynomial is related to the geometric dimension $V(I')$, at least for a monomial ideal I'.

Some thought reveals that $\mathrm{HP}_{R'}$ grows at least linearly iff at least one of the coordinate axes has no black dots. Assuming, e.g., that all dots on the α_1-axis are white, this means that every generator in the monomial ideal I' is a multiple of one of $x_2, \ldots, x_n$, and hence the x_1-axis is contained in $V(I')$.

Similarly, $\deg \mathrm{HP}_{R'} \geq 2$ iff there is a two-dimensional coordinate plane without a black point. Assuming it is the $\alpha_1\alpha_2$ plane, we can see that the x_1x_2-plane is contained in $V(I')$, and so on—in general, the degree of the Hilbert polynomial is the largest dimension of a coordinate subspace contained in $V(I')$. (And since I' is a monomial ideal, $V(I')$ is the union of coordinate subspaces.)

The proofs relating the Hilbert polynomial to the other definitions of dimension and degree mentioned earlier are not too difficult, but here we do not treat them.

8.4 Computation with ideals and Gröbner bases

Here we briefly consider algorithmic questions concerning varieties and ideals.

A basic question is *ideal membership.* Given an ideal $I \subseteq \Bbbk[x_1, \ldots, x_n]$, specified by a list of generators, i.e., $I = \langle f_1, \ldots, f_m \rangle$, how can we test whether a given polynomial g belongs to I?

Recalling that $g \in I$ means $g = \sum_{i=1}^m h_i f_i$ for some h_i, one way might be to look for the h_i, say by solving a system of linear equations for their coefficients. But, as we have remarked, the required degrees of the h_i may be very high, and this method is not practical.

If we have $n = 1$, i.e., univariate polynomials, every ideal can be generated by a single polynomial f, and testing whether $g \in \langle f \rangle$ is very simple: we just reduce g modulo f and see if the remainder is 0. This, of course, assumes that we know a single generator: if I is given by several generators $f_1, \ldots, f_m$, then we first need to compute their greatest common divisor.

The division algorithm. Back in the multivariate setting and trying to proceed analogously, the first question is, given generators $f_1, \ldots, f_m$, what does it mean to reduce them "modulo $f_1, \ldots, f_m$"? We would like to write $g = a_1 f_1 + \cdots + a_m f_m + r$, for suitable polynomials $a_1, \ldots, a_m$ and r, where r should be a "remainder" after the division of g by the f_i.

A good way of doing this is to fix a monomial ordering $\leq$, as introduced in the previous section (but this time it need not be graded), and always try to get rid of the leading monomial of the current g by subtracting the right multiple of some f_i.

Here is the division algorithm. It receives g as input, and successively reduces it by subtracting suitable multiples of the f_i, while simultaneously building the remainder r.

1. Set $r := 0$.

2. Let $\mu := \mathrm{LM}(g)$ be the leading monomial of the current g. If there is some i such that $\mathrm{LM}(f_i)$ divides μ, choose one (arbitrarily), and subtract the appropriate multiple of f_i from g so that the coefficient of μ after the subtraction is 0. Repeat this step with the new g. If there is no such i, go to the next step.

3. At this point none of the $\mathrm{LM}(f_i)$ divides $\mathrm{LM}(g)$. Subtract the leading term of g (i.e., $\mathrm{LM}(g)$ with the coefficient it has in g) from g and add it to r. If $g = 0$, finish, and otherwise, go back to the previous step.

This algorithm is finite, since it strictly decreases $\mathrm{LM}(g)$, according to the monomial ordering, in each step.

But unfortunately, it is not sufficient to test ideal membership unless we have a very good set of generators. For example, if we run it with $f_1 = x^2 + y$, $f_2 = xy + x$, $g = x^2 - y^2$, and the graded lexicographic order as in the preceding section, we get a nonzero remainder $-y^2 - y$. Yet $g \in \langle f_1, f_2 \rangle$, since $x^2 - y^2 = -yf_1 + xf_2$.

The problem here is that, in the expression $-yf_1 + xf_2$, the leading terms cancel out.

Gröbner bases. It turns out that, for a given monomial ordering, every polynomial ideal I has a "very good" set of generators, called Gröbner[5] basis, for which the division algorithm above is guaranteed to test membership in I correctly: it returns remainder 0 iff $g \in I$.

This can be taken as a definition of a Gröbner basis. An equivalent condition, and the usual definition, is this:

An m-tuple $f_1, \ldots, f_m$ is a **Gröbner basis** of an ideal I, w.r.t. a given monomial order, if $I = \langle f_1, \ldots, f_m \rangle$ and $\mathrm{LM}(I) = \langle \mathrm{LM}(f_1), \ldots, \mathrm{LM}(f_m) \rangle$; in words, the leading monomials of the f_i generate the ideal of the leading monomials of all polynomials in I.

A Gröbner basis of I w.r.t. one monomial order may fail to be a Gröbner basis for a different monomial order.

[5] Often spelled Groebner in English texts and software.

A Gröbner basis $f_1, \dots, f_m$ is called *reduced* if it satisfies a certain natural minimality condition, namely, the leading monomials of the f_i have coefficient 1, and no monomial in any f_i is in the ideal generated by the $\mathrm{LM}(f_j)$ for $j \neq i$. For a given I and monomial order, it can be shown that a reduced Gröbner basis is unique.

There are algorithms that, given an arbitrary set of generators of I, compute a Gröbner basis, usually a reduced one, w.r.t. a given monomial order. This algorithmic task has been investigated a lot, since it is very significant both in theory and in practice. In the worst case, the computational complexity, as well as the size of the resulting Gröbner basis, are at least exponential in n, the number of variables.

Once a Gröbner basis is available, we can solve the ideal membership problem by the division algorithm. Many other tasks can be solved as well: computing the sum, intersection, or quotient of two ideals; computing the dimension, Hilbert polynomial, and Hilbert function of a given variety; solving a system of polynomial equations; etc. The worst-case computational complexity of these problems is again very high, but the existing implementations can sometimes handle impressively large instances.

A nice mathematical application of these algorithms is for automatic theorem proving: with Gröbner bases and some cleverness one can make a computer program routinely prove many theorems in high-school geometry or even beyond it, for example, the Pappus theorem. The method is sketched in [CLO07].

Here we finish our very brief excursion to algorithms, referring to [CLO07] for a thorough introduction.

8.5 Projective varieties

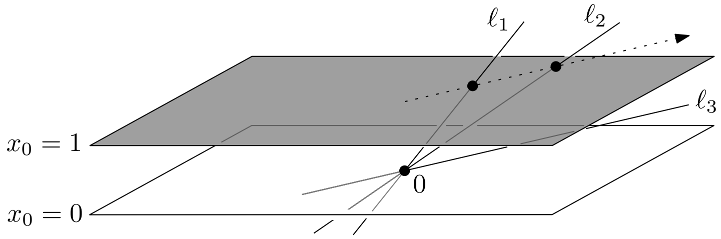
Instead of the affine space $\mathbb{k}^n$, algebraic geometry is usually done in the projective space $\mathbb{P}^n = \mathbb{P}^n(\mathbb{k})$, which can be thought of as a completion of $\mathbb{k}^n$ by adding points at infinity in a suitable way. Then almost everything comes out more elegantly and algebraic varieties behave much better—for example, over an algebraically closed field, the projection of a variety is again a variety, unlike in the affine case.

The projective space. To construct $\mathbb{P}^n$ formally, we consider all $(n+1)$-tuples $(a_0 : a_1 : \cdots : a_n)$, where $a_0, \dots, a_n \in \mathbb{k}$ are not all simultaneously 0. Each point a of $\mathbb{P}^n$ is an equivalence class of such $(n+1)$-tuples consisting of all nonzero multiples of some $(n+1)$-tuple:

$$a = \{(\lambda a_0 : \lambda a_1 : \cdots : \lambda a_n) : \lambda \in \mathbb{k} \setminus \{0\}\}.$$

Such an equivalence class can be viewed as a line through the origin in $\mathbb{k}^{n+1}$. The $(n+1)$-tuple $(a_0 : \cdots : a_n)$ is called the **homogeneous coordinates** of a; these are defined only up to a scalar multiple.

The following picture illustrates, for the case $n = 2$, the geometric meaning of this construction.



Here $\Bbbk^2$, to which we want to add points at infinity, is embedded in $\Bbbk^3$ as the gray plane $x_0 = 1$ (where the coordinates in $\Bbbk^3$ are x_0, x_1, x_2 and the x_0-axis is drawn vertical). Each point a of this plane corresponds to the line $\overline{0a}$ through the origin in $\Bbbk^3$.

Conversely, each line through the origin corresponds to exactly one point of the gray plane, *except* for horizontal lines, such as ℓ_3. When we start tilting the line ℓ_1 towards the position ℓ_2 and further towards the horizontal position ℓ_3, the corresponding point in the gray plane recedes to infinity along the dashed line. So horizontal lines such as ℓ_3 correspond to points at infinity, one point for each direction of parallel lines in the gray plane.

Algebraically, in this interpretation, a point of $\mathbb{P}^n$ with homogeneous coordinates $(x_0 : \cdots : x_n)$ with $x_0 \neq 0$ corresponds to the point $(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0}) \in \Bbbk^n$. Adding the points at infinity, with $x_0 = 0$, can be thought of as adding a copy of $\mathbb{P}^{n-1}$ to $\Bbbk^n$.

On the other hand, the structure of $\mathbb{P}^n$ is the same everywhere and locally, in the neighborhood of each point, it looks like the affine space $\Bbbk^n$. In our picture, the plane representing $\Bbbk^2$ can be rotated around 0, and this yields various ways of placing $\Bbbk^2$ in $\mathbb{P}^2$. In algebraic geometry, this allows one to transfer all kinds of "local" notions from the affine setting to the projective one.

Projective varieties. We would like to say that projective varieties are zero sets of polynomial systems of equations in $\mathbb{P}^n$, but we have to be a bit careful.

Working in $\mathbb{P}^n$, we have $n+1$ coordinates $x_0, \ldots, x_n$, but it does not make sense to consider, for example, the equation $x_1 = x_0^2$, since the $(n+1)$-tuple $(1 : 1 : \cdots : 1)$ satisfies it, but $(2 : 2 : \cdots : 2)$, representing the same point of $\mathbb{P}^n$, does not.

One has to consider only zero sets of **homogeneous polynomials** $f \in \Bbbk[x_0, \ldots, x_n]$, meaning that all monomials of f have the same degree; then the zero set can be regarded as a subset of $\mathbb{P}^n$. The counterpart for ideals is a **homogeneous ideal**, one generated by homogeneous polynomials (but necessarily containing non-homogenerous polynomials too); for such an ideal I, the variety $V(I) \subseteq \mathbb{P}^n$ is well defined as the set of common zeros of all $f \in I$.

Every polynomial $f \in \Bbbk[x_1, \ldots, x_n]$ can be *homogenized* to a homogeneous polynomial $\tilde{f}$ by adding an appropriate power of x_0 to each term so that the resulting polynomial becomes homogeneous (and has the same degree as f). For instance, from $x_1^3+x_1x_2+5$ we get $x_1^3+x_0x_1x_2+5x_0^3$. An ideal $I \subseteq \Bbbk[x_1, \ldots, x_n]$ is homogenized to the homogeneous ideal $\tilde{I} = \langle \tilde{f} : f \in I \rangle \subset \Bbbk[x_0, \ldots, x_n]$. (Let us mention that the homogenization of a generating set of I need not generate $\tilde{I}$.) From an affine variety $V(I) \subseteq \Bbbk^n$ we thus obtain the **projective completion** $V(\tilde{I}) \subseteq \mathbb{P}^n$ (it is perhaps worth mentioning that isomorphic affine varieties may have nonisomorphic projective completions). The meaning of $I(X)$ for $X \subseteq \mathbb{P}^n$ is also modified appropriately.

Many of the concepts and results from the affine setting transfer to projective varieties without change (irreducible decomposition, Zariski open and closed sets) or with only minor modifications.

For the weak Nullstellensatz, $V(I) = \emptyset$ not only for $I = \langle 1 \rangle$, but also when the radical of I is $\langle x_0, x_1, \ldots, x_n \rangle$. This *irrelevant ideal* also has to be excluded in the strong Nullstellensatz; after that, over an algebraically closed field, we have a bijective correspondence between homogeneous radical ideals and projective varieties.

A **morphism** $f \colon X \to Y$ of projective varieties $X \subseteq \mathbb{P}^m$ and $Y \subseteq \mathbb{P}^n$ needs to be defined locally: for each $x_0 \in X \subseteq \mathbb{P}^m$ there is a Zariski open neighborhood U and homogeneous polynomials $f_0, \ldots, f_n \in \Bbbk[x_0, \ldots, x_m]$ of the same degree such that $f(x) = (f_0(x) : \cdots : f_n(x))$ for all $x \in U$ (and in particular, at least one $f_i(x)$ must be nonzero for each x).

As for the Hilbert function, in the projective case one needs to take the dimension of $\Bbbk[x_0, \ldots, x_n]/I_{=d}$, where $I_{=d}$ is the vector subspace spanned by homogeneous polynomials of degree *exactly* d in the homogeneous ideal I.

Cutting with a polynomial. If X is a k-dimensional projective variety over an algebraically closed field and f is a polynomial, then $k - 1 \leq \dim(X \cap V(f)) \leq k$. If, moreover, X is irreducible and f does not vanish on it, then $\dim(X \cap V(f)) = k - 1$.

Exercise 8.8. *Show that this fails for affine varieties;* $\dim(X \cap V(f))$ *can be smaller than* $\dim X - 1$.

Projection. Unlike in the affine case, the projection of a projective variety is also a projective variety, and so is the image under a morphism.

One has to be slightly careful with what is meant by a projection, since in $\mathbb{P}^n$ we cannot simply omit some of the homogeneous coorditates, because we might get all 0s.

One way around this is to consider a projection as a map $\pi \colon \mathbb{P}^m \times \mathbb{P}^n \to \mathbb{P}^n$, but strictly speaking, we have not yet defined what a variety in $\mathbb{P}^m \times \mathbb{P}^n$ is. There are two equivalent ways of doing that.

First, we may embed $\mathbb{P}^m \times \mathbb{P}^n$ as a variety in $\mathbb{P}^{(m+1)(n+1)-1}$; this is called the **Segre embedding**, and it sends a pair $((x_0 : \cdots : x_m), (y_0 : \cdots : y_n))$ to $(x_0 y_0 : x_1 y_0 : \cdots : x_i y_j : \cdots : x_m y_n)$. Then varieties in $\mathbb{P}^m \times \mathbb{P}^n$ are just the intersections of varieties in $\mathbb{P}^{(m+1)(n+1)-1}$ with the embedded copy of $\mathbb{P}^m \times \mathbb{P}^n$. (We note in passing that the image of the Segre embedding is essentially the determinantal variety $D_2(m+1, n+1)$ mentioned in Section 8.2.)

Second, and more explicitly, a variety in $\mathbb{P}^m \times \mathbb{P}^n$ is the common zero set of a set of *bihomogeneous* polynomials $f \in \Bbbk[x_0, \ldots, x_m, y_0, \ldots, y_n]$, where f is bihomogeneous if each monomial has degree k in the x_i and degree ℓ in the y_i, for some k, ℓ, possibly with $k \neq \ell$. Then the result can be stated as follows:

Theorem 8.9 (Projection theorem)**.** *For every projective variety* $Z \subseteq \mathbb{P}^m \times \mathbb{P}^n$ *over an algebraically closed field,* $\pi(Z)$ *is also a projective variety, where* $\pi \colon \mathbb{P}^m \times \mathbb{P}^n \to \mathbb{P}^n$ *is the projection onto the second factor.*

Let us prove at least something in this long section.

Proof. Let $f_1, \dots, f_r$ be bihomogeneous generators of $I(Z)$. We may assume that all of them have the same degree k in the x_i. (If, in order to achieve that, we need to raise the degree of f_1 by d, say, we replace f_1 by the $(n+1)$-tuple of polynomials $x_0^d f_1, x_1^d f_1, \dots, x_n^d f_1$, which does not change the zero set.)

Let us fix a point $a \in \mathbb{P}^n$ and write $f_{i,a}(x) := f_i(x,a)$. By definition, $a \notin \pi(Z)$ means that the $f_{i,a}$ have no common zero in $\mathbb{P}^m$. By the projective weak Nullstellensatz, this happens iff the radical of the homogeneous ideal $I := \langle f_{1,a}, \dots, f_{r,a} \rangle \subseteq \Bbbk[x_0, \dots, x_n]$ contains the irrelevant ideal $\langle x_0, \dots, x_n \rangle$. In other words, there are $s_0, \dots, s_n$ with $x_i^{s_i} \in I$, $i = 1, 2, \dots, n$.

For the proof to work, we transform this condition further: setting $s := s_0 + \cdots + s_n$, we can see that I contains *all* homogeneous polynomials of degree s. So $a \notin \pi(X)$ if and only if there exists s such that for every homogeneous $g \in \Bbbk[x_0, \dots, x_n]$ of degree s we can find $h_1, \dots, h_r \in \Bbbk[x_0, \dots, x_n]$ with

$$g = \sum_{i=1}^{r} h_i f_{i,a}. \tag{3}$$

Here, crucially, since all the $f_{i,a}$ are homogeneous of degree k, we may assume that the h_i are homogeneous of degree $s-k$, because monomials of any other degree can be discarded from them without changing the validity of (3).

Therefore, for every g, (3) can be rewritten as a system of linear equations for the unknown coefficients of the h_i. The matrix of this system, call it A, does not depend on g, and its entries are homogeneous polynomials in $a_0, \dots, a_n$, the homogeneous coordinates of a. The number of equations is t, the number of monomials of degree s in $n+1$ variables; it equals $\binom{s+n}{n}$ but we do not need that.

The solvability of (3) for every g means that the linear system is solvable for every right-hand side, which means exactly that A has rank t. Hence the *negation* of this condition can be expressed as vanishing of all the $t \times t$ minors of A.

Let Y_s be the set of all $a \in \mathbb{P}^n$ such that the matrix A as above has rank less than t. Each Y_s is a variety, and we have $\pi(Z) = \bigcap_{s=0}^{\infty} Y_s$. Therefore, $\pi(Z)$ is a projective variety as claimed. □

9 Bézout's inequality in higher dimensions

9.1 In search of a proper statement

We again consider the system of polynomial equations $f_1 = 0, \dots, f_m = 0$, $f_1, \dots, f_m \in \Bbbk[x_1, \dots, x_n]$, this time for $n > 2$ variables.

The most important case is $m = n$. Guided by the example with hyperplanes, i.e., with $f_i = (x_i - 1) \cdots (x_i - d_i)$, where $d_i = \deg f_i \geq 1$, we expect that if the number of solutions is finite, then it should be at most $d_1 d_2 \cdots d_n$. Moreover, finitely many solutions should be the typical, "generic" case.

Warning example. Unlike in the planar case, over an arbitrary field, having finitely many solutions does *not* guarantee that the bound $d_1 d_2 \cdots d_n$ for their number is correct.

Indeed, the system of three equations

$$(x-1)^2(x-2)^2 \cdots (x-k)^2 + (y-1)^2(y-2)^2 \cdots (y-k)^2 = 0,\ z=0,\ z=0$$

has k^2 solutions in $\mathbb{R}^3$, but the degrees are $2k, 1, 1$. We note that the solution set in $\mathbb{C}^3$ is infinite.

Another example. In the previous example, the first equation has only 1-dimensional solution set over $\mathbb{R}$, while the two remaining equations are identical. However, over $\mathbb{C}$ the solution set of the first equation is 2-dimensional, and so at least over algebraically closed fields, one might hope to exclude this kind of pathology by imposing a suitable condition on the f_i. Indeed, drawing inspiration from the planar case, a natural guess for such a condition can be that no two of the f_i have a common factor.

However, things are not that simple, and the suggested condition is definitely not the right one. Here is a highly instructive example for $n = 3$:

$$f_1 = x^3 - yz,\ f_2 = y^2 - xz,\ f_3 = z^2 - x^2 y.$$

These are irreducible polynomials, as is easy to check, none a multiple of another. But $V(f_1, f_2, f_3)$ contains the curve C with parametric expression

$$C = \{(t^3, t^4, t^5) : t \in \mathbb{C}\},$$

and so surely it is not finite.

This example is also interesting in another respect. In linear algebra, every k-dimensional vector subspace of $\Bbbk^n$ can be described by $n-k$ linear equations; for example, a line in $\mathbb{R}^3$ is always the intersection of two planes. In contrast, the curve C *cannot* be defined by two polynomial equations: It is easy to check the common zero set of every two of the f_i contains points not belonging to the zero set of the third—e.g., $V(f_1, f_2)$ contains the z-axis, where f_3 is nonzero. With more effort, one can show that *no two* polynomials suffice; this is done algebraically, by checking that the ideal $\langle f_1, f_2, f_3 \rangle$ cannot be generated by two polynomials.

Let us remark that things cannot get completely out of hands with this kind of examples: it is known that every irreducible affine variety in $\Bbbk^n$, $\Bbbk$ algebraically closed, can be given as the zero set of at most $n+1$ polynomials [Hei83, Prop. 3].

Bézout's inequality assuming finitely many zeros. It seems that there is no particularly useful general condition for $V(f_1, \ldots, f_n)$ to be finite, although there are algorithms that can decide this question for any given $f_1, \ldots, f_n$—but these are nontrivial and quite demanding computationally.

One way around this is to *assume* $V(f_1, \ldots, f_n)$ finite. Then, for $\Bbbk$ is algebraically closed, the expected inequality for the number of zeros does hold.

Theorem 9.1 (Higher-dimensional Bézout's inequality I)**.** *Let $\Bbbk$ be algebraically closed, and let $f_1, \ldots, f_n \in \Bbbk[x_1, \ldots, x_n]$ be polynomials of degrees $d_1, \ldots, d_n \geq 1$. Assuming that $V(f_1, \ldots, f_n) \subset \Bbbk^n$ is* finite*, it has at most $d_1 d_2 \cdots d_n$ points.*

Actually, one can say a bit more: even if $V(f_1,\dots,f_n)$ contains irreducible components of positive dimension, the number of one-point irreducible components is still at most $d_1d_2\cdots d_n$.

We will not prove Theorem 9.1 here. A reasonably accessible algebraic proof can be found in [Tao12, Sec. 8.4].

Bounding the number of nonsingular zeros. The above formulation of Bézout's inequality leaves something to be desired, since, as we have mentioned, verifying the assumption $|V(f_1,\dots,f_n)|<\infty$ is not easy in general (although there are various sufficient conditions known; see, e.g., [CLO05, Chap. 3,4] and [Sch95]).

Another formulation, which is often useful for applications, is to consider only a suitable kind of "nice" zeros, namely, only those where the hypersurfaces $X_i := V(f_i)$ intersect transversally..

We will work only over the field $\mathbb{R}$, where one can rely on intuition and methods from analysis. However, with an appropriate generalization of notions like gradient, results can also be obtained for other fields—see [CKW11, Sec. 5].

Let $X_1,\dots,X_n\subseteq\mathbb{R}^n$ be the hypersurfaces as above and let a be a point where they all intersect. Transversality means that if we make a tangent hyperplane h_i to each X_i at a, then these n hyperplanes intersect only in a—they look like the coordinate hyperplanes, after a suitable affine transformation (this includes the assumption that each X_i is $(n-1)$-dimensional in some neighborhood of a).

We recall that if $f\colon\mathbb{R}^n\to\mathbb{R}$ is a differentiable function, then the gradient ∇f at a point a is the "fastest ascent" direction for f. Assuming $f(a)=0$, $\nabla f(a)$ is perpendicular to the zero set of f, and thus it is a normal vector of the tangent hyperplane of the zero set, assuming $\nabla f(a)\neq 0$. (Rigorously this can be derived from the implicit function theorem.)

The transversality of our $X_1,\dots,X_n$ at a thus corresponds to linear independence of the n gradients $\nabla f_1(a),\dots,\ \nabla f_n(a)$, or in other words, to the **Jacobian determinant**

$$J_{f_1,\dots,f_n}(a) := \det\begin{pmatrix}\frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a)\\ \vdots & \dots & \vdots\\ \frac{\partial f_n}{\partial x_1}(a) & \dots & \frac{\partial f_n}{\partial x_n}(a)\end{pmatrix}$$

being nonzero. (Apologies to the readers for whom the geometric meaning of the Jacobian is well known and boring.)

A point $a\in V(f_1,\dots,f_n)$ with $J_{f_1,\dots,f_n}(a)\neq 0$ is called a **nonsingular zero**.

Theorem 9.2 (Higher-dimensional Bézout's inequality II)**.** *Let $f_1,\dots,f_n\in\mathbb{R}[x_1,\dots,x_n]$. Then the polynomial system $f_1=0,\dots,\ f_n=0$ has at most $d_1d_2\cdots d_n$ nonsingular zeros in $\mathbb{R}^n$, where $d_i=\deg f_i$.*

9.2 Proof for nonsingular zeros

We present a proof of Theorem 9.2 due to Wooley [Woo96], mostly following a presentation in [CKW11, Sec. 5]. For another more or less elementary proof, going via a complex version of the theorem, see [BPR03, Sec. 4.7].

So let $a_1, \ldots, a_N \in \mathbb{R}^n$ be nonsingular common zeros of $f_1, \ldots, f_n$; we want to show that $N \leq D = d_1 d_2 \cdots d_n$.

First we fix a linear polynomial $\pi \in \mathbb{R}[x_1, \ldots, x_n]$ such that the $\pi(a_i)$ are all distinct; we can think of this as choosing a projection on a suitable line. Armed with the knowledge from the previous sections, the reader will surely supply a rigorous proof of existence of a suitable π.

The general idea of the proof is to produce a nonzero univariate polynomial of degree at most D for which all the $\pi(a_i)$ are roots.

To this end, we would like to have a polynomial $h \in \mathbb{R}[y_1, \ldots, y_n, z]$ satisfying the following conditions:

(C1) The polynomial $\tilde{h} := h(f_1, \ldots, f_n, \pi) \in \mathbb{R}[x_1, \ldots, x_n]$, obtained by substituting $f_i(x_1, \ldots, x_n)$ for y_i and $\pi(x_1, \ldots, x_n)$ for z into h, is the zero polynomial.

(C2) The highest power of z occurring in h is at most D.

(C3) The univariate polynomial $h_0(z) := h(0, 0, \ldots, 0, z)$ is nonzero.

If we had such an h, we would be done: by (C1), $h_0(\pi(a)) = 0$ whenever a is a common zero of the f_i, by (C2) we have $\deg h_0 \leq D$, and together with (C3) this would show that h_0 has at most D zeros.

However, a suspicious thing is that this plan does not use the nonsingularity of the considered common zeros of the f_i, and indeed, we will have to modify it. But (C1) and (C2) can be achieved; this is done by linear algebra and counting, and it works over any field.

Lemma 9.3. *Given arbitrary polynomials $f_1, \ldots, f_n, \pi \in \Bbbk[x_1, \ldots, x_n]$ with $\deg f_i = d_i$ and $\deg \pi = 1$, there exists a nonzero polynomial $h \in \Bbbk[y_1, \ldots, y_n, z]$ satisfying (C1) and (C2).*

We postpone the proof of the lemma. Having such an h, we cannot guarantee (C3), unfortunately. But here we use the assumption with nonsingular zeros to perturb the f_i, and for the perturbed version we will be able to get (C3).

Concretely, we perturb by choosing a sufficiently small vector $\delta = (\delta_1, \ldots, \delta_n) \in \mathbb{R}^n$ and considering the perturbed system $f_1 = \delta_1, \ldots, f_n = \delta_n$. We claim that if a is a nonsingular zero of the original system, with zero right-hand sides, then for every δ sufficiently small, there is a zero $a(\delta)$ of the perturbed system, such that $a(\delta) \to a$ as $\|\delta\| \to 0$.

This is a textbook application of the implicit function theorem; after all, nonzero Jacobian is typically used through that theorem. We just consider the function $F \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ given by coordinate-wise by $F(x, \delta)_i := f_i(x) - \delta_i$. Then $F(a, 0) = 0$, and the implicit function theorem guarantees the existence of a (continuous) function $a(\delta)$ with $F(a(\delta), \delta) = 0$ for all δ sufficiently small (note that the Jacobian in the implicit function theorem is with respect to the "dependent" variables, which in our case are the x_i, and this is exactly $J_{f_1,\ldots,f_n}(a)$ as in the definition of nonsingular zero).

It follows that if the original system has at least N nonsingular zeros, so does the perturbed system for δ sufficiently small. Moreover, again for δ small enough, these N zeros of the perturbed system still yield N distinct values of the projection π. So if h satisfies (C1) and (C2), then for every $\delta \in \mathbb{R}^n$ sufficiently small, $h(\delta_1, \ldots, \delta_n, z)$ vanishes for at least N distinct values of z.

At the same time, since $V(h)$ has zero Lebesgue measure (or, alternatively, by the Schwartz–Zippel theorem), there are values $\bar{\delta}_1, \ldots, \bar{\delta}_n \in (-\delta, \delta)$ and $\bar{z} \in \mathbb{R}$ with $h(\bar{\delta}_1, \ldots, \bar{\delta}_n, \bar{z}) \neq 0$. It follows that $h(\bar{\delta}_1, \ldots, \bar{\delta}_n, z)$ is a nonzero polynomial in z, of degree at most D by (C2), and hence $N \leq D$ as claimed. It remains to prove the lemma.

Proof of Lemma 9.3. We will look for h in the form

$$h(y_1, \ldots, y_n, z) = \sum_{\alpha \in A} c_\alpha y_1^{\alpha_1} \cdots y_n^{\alpha_n} z^{\alpha_{n+1}},$$

where $A \subset \mathbb{Z}_{\geq 0}^{n+1}$ is a suitable finite set of $(n+1)$-tuples, whose choice we will discuss later, and where the c_α are regarded as unknowns. So we have $|A|$ unknowns.

If we make the substitution $y_1 = f_1$,..., $y_n = f_n$, $z = \pi$ for a monomial $y_1^{\alpha_1} \cdots y_n^{\alpha_n} z^{\alpha_{n+1}}$, the degree of the resulting polynomial in $x_1, \ldots, x_n$ is

$$d_1\alpha_1 + \cdots + d_n\alpha_n + \alpha_{n+1}.$$

Let us call this expression the *weight* $w(\alpha)$, and set $w(A) := \max_{\alpha \in A} w(\alpha)$.

Thus, if we fix A, the degree of $\tilde{h}$, the polynomial after the substitution, is at most $w(A)$. Moreover, the coefficients of $\tilde{h}$ are *linear* functions of the c_α.

We want to force zero coefficient for every monomial that could possibly appear in $\tilde{h}$; each such requirement yields a linear equation for the c_α. Since $\deg \tilde{h} \leq w(A)$, we thus obtain $\binom{w(A)+n}{n}$ homogeneous linear equations for $|A|$ unknowns.

Hence the lemma will be proved as soon as we find A such that $|A| > \binom{w(A)+n}{n}$ and $\alpha_{n+1} \leq D$ for all $\alpha \in A$.

For an integer W, let

$$A = A(W) := \{\alpha : w(\alpha) \leq W, \alpha_{n+1} \leq D\}.$$

We want to show that $|A(W)| > \binom{W+n}{n}$ holds for all sufficiently large W. The counting must be quite precise; after all, the proof cannot work with $D-1$ instead of D.

For a parameter T, let $N(T)$ be the number vectors $(\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_{\geq 0}^n$ such that $\sum_{i=1}^n d_i\alpha_i \leq T$; we have

$$|A(W)| = \sum_{\alpha_{n+1}=0}^{D} N(W - \alpha_{n+1}) \geq (D+1)N(W-D).$$

Let $B = B(T)$ be the set of all $\beta = (\beta_1, \ldots, \beta_n) \in \mathbb{Z}_{\geq 0}^n$ with $\beta_1 + \cdots + \beta_n \leq T$; we have $|B| = \binom{T+n}{n}$. We can express $N(T)$ as the number of $\beta \in B$ such that $\beta_i \bmod d_i = 0$ for all i.

Let $r(\beta) = (\beta_1 \bmod d_1, \ldots, \beta_n \bmod d_n)$, and let us partition B into equivalence classes according to the value of $r(\beta)$; there are $d_1 d_2 \cdots d_n = D$ classes. It is easy to see that the class with $r(\beta) = (0, \ldots, 0)$ is at least as large as any other class, and so

$$N(T) \geq \frac{1}{D}\binom{T+n}{n}.$$

Consequently,

$$\begin{aligned} |A(W)| &\geq (D+1)N(W-D) \geq \frac{D+1}{D}\binom{W-D+n}{n} \\ &= \frac{D+1}{D}\binom{W+n}{n} \cdot \frac{(W-D+n)\cdots(W-D+1)}{(W+n)\cdots(W+1)} \\ &\geq \binom{W+n}{n}\frac{D+1}{D}\left(1-\frac{D}{W+1}\right)^n. \end{aligned}$$

For D fixed and $W \to \infty$, we have $(1-\frac{D}{W+1})^n \to 1$, while $\frac{D+1}{D}$ remains bounded away from 1. Hence $|A(W)| > \binom{W+n}{n}$ for W sufficiently large as desired. The lemma, as well as Bézout's inequality for nonsingular zeros, are proved. □

10 Bounding the number of connected components

How complicated can the zero set of a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ of degree d be? The answer depends, of course, on how we measure the complexity, and there are several sensible ways.

We will first look at the number of connected components of the complement, i.e., of $\mathbb{R}^n \setminus V(f)$. In this case a good answer can be given with a reasonably simple proof.

To see what can be expected, we consider the usual example with hyperplanes, slightly modified:

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n}\prod_{j=1}^{m}(x_i - j).$$

The degree is $d = mn$, and the zero set, a grid of hyperplanes, partitions $\mathbb{R}^n$ into $(m+1)^n \sim (d/n)^n$ components (axis-parallel boxes).

Theorem 10.1. *For a polynomial $f \in \mathbb{R}[x_1, \ldots, x_n]$ of degree $d \geq 2$, $\mathbb{R}^n \setminus V(f)$ has at most $(d+1)^n$ components.*

The proof below is similar to one in [ST12, Appendix A]. This kind of arguments goes back to Oleinik and Petrovskiĭ [OP49, Ole51], Milnor [Mil64], and Thom [Tho65].

In the proof, we will need the following result.

Fact 10.2. *Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be a polynomial map (that is, a map for which each coordinate $f_i\colon \mathbb{R}^n \to \mathbb{R}$ is given by a polynomial; in algebraic geometry, one usually speaks of* regular maps *in this context), and let $X \subset \mathbb{R}^n$ be a proper algebraic variety (that is, X is contained in the zero set of a nonzero polynomial). Then the image $f(X)$ does not fill any open ball in $\mathbb{R}^n$.*

This result may look obvious, but obvious approaches to proofs have their caveats.

First, we know that X is "small"; e.g., it does not fill any open ball. But, for example, the image of a segment under a continuous map may be a unit square, as is witnessed by the famous *Peano curve*. So we have to use other properties of f besides continuity.

Approaching from the side of mathematical analysis, we can use the fact (which we do not prove here) that the image Lebesgue null set under a smooth map is Lebesgue null, plus Exercise 2.3. In our case, a polynomial map is not only smooth (inifinitely differentiable), but also *locally Lipschitz*, which allows for a quite straightforward proof.

Exercise 10.3. (a) *Verify that a polynomial map $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz, meaning that for every $x_0 \in \mathbb{R}^n$ there exist $\varepsilon > 0$ and L such that f is L-Lipschitz on the ε-ball around x_0, i.e., $\|f(x) - f(y)\| \le L\|x - y\|$ for every choice of x, y in that ball. (Unlike in most uses of the letter ε in analysis, here one can actually take ε as large as desired.)*

(b) *Prove that the image of a Lebesgue null set under a locally Lipschitz map $\mathbb{R}^n \to \mathbb{R}^n$ is Lebesgue null.*

A more algebraic approach to Fact 10.2 would be to prove that the image of a proper subvariety in $\mathbb{R}^n$ under a polynomial map is a proper subvariety of $\mathbb{R}^n$. Unfortunately, this is not literally true, as can be seen by modifying the hyperbola example from Section 8.2. What can be shown is that such an image is *contained* in a proper subvariety of $\mathbb{R}^n$, which is enough for our purposes. This is not too hard, given the tools covered so far, and it is a special case of a result stating that a regular map cannot increase dimension, but here we will not go through the argument.

Proof of Theorem 10.1. First we count only the *bounded* components of $\mathbb{R}^n \setminus V(f)$.

We do not know apriori that there are only finitely many components, but for some of the arguments below it will be important that we work with finitely many. So we fix any collection $\mathcal{C}$ of finitely many bounded components of $\mathbb{R}^n \setminus V(f)$ and work only with these. We will show that $|\mathcal{C}| \le (d-1)^n$, which will imply, in particular, that there are only finitely many components altogether.

For each component $C \in \mathcal{C}$, we have either $f > 0$ or $f < 0$ on C; let us assume the former. We claim that f attains at least one maximum on C. Indeed, f attains some positive value $\varepsilon > 0$ at some point of C, the set $\{x \in C : f(x) \ge \frac{\varepsilon}{2}\}$ is compact and nonempty, and so f attains a maximum at some x_C there.

Since x_C lies inside the open set C and f is differentiable, the gradient ∇f vanishes at x_C, and hence $x_C \in V(\nabla f)$, where $V(\nabla f)$ is a shorthand for the set of common zeros of $\frac{\partial f}{\partial x_i}$, $i = 1, 2, \ldots, n$.

We note that deg $\frac{\partial f}{\partial x_i} \le d - 1$. The idea is to apply Bézout's inequality, in the form with nonsingular zeros, to bound $|V(\nabla f)|$, and hence the number of bounded components, by $(d-1)^n$.

The condition for nonsingularity of a common zero a of the $\frac{\partial f}{\partial x_i}$ reads $\det H_f(a) \neq 0$, where H_f is the *Hessian matrix* of f, with

$$(H_f)_{ij} := \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

However, we cannot guarantee that $\det H_f$ is not identically 0 (even some of the partial derivatives may be identically 0—for example, if f does not depend on some of the variables).

The next trick is to perturb the function whose maxima we seek. Indeed, if the maximum of f over a bounded component C is at least ε, then another function $\tilde{f}$ differing from f by at most $\frac{\varepsilon}{3}$, say, also has to attain a maximum in C. (Note that $C \in \mathcal{C}$ is still one of the original components of $\mathbb{R}^n \setminus V(f)$, even though we maximize the perturbed function $\tilde{f}$ over it.)

We actually make two perturbations. First, for δ sufficiently small, we set $\tilde{f} := f - \delta(x_1^2 + \cdots + x_n^2)$. This is the simplest kind of perturbation that may make the Hessian determinant nonzero (if we were willing to use Theorem 9.1 instead of Theorem 9.2, we could skip this perturbation).

It is easy to see that $H_{\tilde{f}} = H_f - 2\delta I$, where I is the identity matrix, and hence $\det H_{\tilde{f}} = 0$ exactly if 2δ is an eigenvalue of H_f. Thus, no matter what H_f looks like, $\det H_{\tilde{f}}$ is a nonzero polynomial for all but finitely many δ. We fix some sufficiently small δ for which $\det H_{\tilde{f}}$ is not identically zero; then $\tilde{f}$ is fixed too.

Next, we let $\tilde{f}_\eta := \tilde{f} - \eta_1 x_1 - \cdots - \eta_n x_n$, where $\eta = (\eta_1, \ldots, \eta_n)$ is a vector of parameters. Then $\nabla \tilde{f}_\eta = \nabla \tilde{f} - \eta$, and so instead of counting the points in $V(\nabla \tilde{f}) = (\nabla \tilde{f})^{-1}(0)$, we now need to count the number of preimages of η under the polynomial map $\nabla \tilde{f} \colon \mathbb{R}^n \to \mathbb{R}^n$. (Geometrically, replacing $\tilde{f}$ with $\tilde{f}_\eta$ corresponds to slightly tilting the originally vertical direction in which we seek maxima or minima of $\tilde{f}$.)

We want to choose η sufficiently small (so that $\tilde{f}_\eta$ and f are sufficiently close) and such the Hessian determinant $\det H_{\tilde{f}_\eta} = \det H_{\tilde{f}}$ does not vanish at the points of the preimage $(\nabla \tilde{f})^{-1}(\eta)$.

The variety of the Hessian determinant, $Y := V(\det H_{\tilde{f}})$, is the zero set of a nonzero polynomial, and $\nabla \tilde{f}$ is a polynomial mapping $\mathbb{R}^n \to \mathbb{R}^n$. Hence by Fact 10.2, there are arbitrarily small η avoiding the image of $\nabla \tilde{f}(Y)$.

For such η, all the maxima and minima of $\tilde{f}_\eta$ are nonsingular common zeros of the polynomials in $\nabla \tilde{f}_\eta$, and so we can bound their number by $(d-1)^n$ as desired.

It remains to account for the unbounded components. For that, we replace f with $g := f \cdot (x_1^2 + \cdots + x_n^n - R^2)$, where R is a sufficiently large number; that is, to the zero set of f we add a large sphere. Then every component of $\mathbb{R}^n \setminus V(f)$ appears as a *bounded* component of $\mathbb{R}^n \setminus V(g)$. Since $\deg g = \deg f + 2$, the bound claimed in the theorem follows. □

A stronger version. The theorem just proved can be strengthened in several respects.

First, there is a quantitative improvement, which becomes significant if the degree d and the dimension n are comparable: the true bound is more like $(d/n)^n$ (which is the lower bound we got from the simple example) than d^n.

Second, the bound can be extended to the complement of the union of several zero sets, i.e., $\mathbb{R}^n \setminus (V(f_1) \cup \cdots \cup V(f_m))$. In this case a reasonably good bound can be obtained just by setting $f = f_1 f_2 \cdots f_m$ and using the bound for a single polynomial.

Third, instead of considering just the complement, which is the set where all of $f_1, \ldots, f_m$ are nonzero, we can consider sets where some of the f_i are required to be 0, some others positive, and some negative. These three improvements are all reflected in the next theorem.

Theorem 10.4. *Let $f_1, \ldots, f_m \in \mathbb{R}[x_1, \ldots, x_n]$ be polynomials of degree at most d, and for every sign vector $\sigma \in \{-1, 0, +1\}^m$ let $S_\sigma \subseteq \mathbb{R}^n$ be defined as*

$$\Big\{x \in \mathbb{R}^n : \operatorname{sgn} f_i(x) = \sigma_i \text{ for all } i = 1, 2, \ldots, m\Big\}.$$

Then, for $m \geq n \geq 2$,

$$\sum_{\sigma \in \{-1,0,+1\}^m} \#S_\sigma \leq \left(\frac{50dm}{n}\right)^n,$$

where $\#S_\sigma$ denotes the number of connected components of S_σ.

The basic ideas of the proof are similar to those in the proof of Theorem 10.1 shown above, but the details are considerably more involved.

In the literature, such results are often stated as bounding the total topological complexity of the considered sets, more precisely, the sum of the *Betti numbers*, instead of just the number of connected components. For still other strengthenings of the just stated theorem, such as a more refined dependence on the degrees of the f_i, as well as replacing the ground set $\mathbb{R}^n$ with a k-dimensional algebraic variety in $\mathbb{R}^n$, see [Bar13] and references therein.

Bounds on the radius of components and inscribed balls. Another way of measuring zero sets of polynomials in $\mathbb{R}^n$ is, for example, by the radius of the smallest ball intersecting all connected components. Here, of course, we need to make some assumptions on the coefficients of the polynomials; typically we assume them to be integers not exceeding some given bound. Here is a general result of this kind:

Theorem 10.5. *Let $f_1, \ldots, f_m \in \mathbb{Z}[x_1, \ldots, x_n]$ be polynomials of maximum degree d whose coefficients are integers bounded by M in absolute value. For $\sigma \in \{-1, 0, 1\}^m$, let $S_\sigma := \{x \in \mathbb{R}^n : \operatorname{sgn} f_i(x) = \sigma_i \text{ for all } i = 1, 2, \ldots, m\}$. Then each connected component of S_σ intersects the ball of radius $R = M^{(d+1)^{Cn}}$ centered at 0, were C is a suitable absolute constant. The bounded connected components of S_σ are all contained in that ball.*

If $\sigma_i \neq 0$ for all i, or in other words, S_σ is defined only by strict inequalities, and if S_σ is nonempty, then it contains a rational point with coordinates whose numerators and denominators are integers not exceeding R in absolute value.

This kind of result goes back to [GV88, Lemma 9] (which deals with more special sets, namely, the zero set of a single polynomial), and the result as above about a ball intersecting all connected components is [BPR96, Theorem 4.1.1] (also see [BPR03, Theorem 13.14]). A statement directly implying the part with the ball containing all bounded components is [BV07, Theorem 6.2]. For the part with a rational point see [BPR03, Theorem 13.15].

On applications. Theorem 10.4 and its relatives have probably hundreds of applications in geometry, combinatorics, computer science and elsewhere. An old but still very beautiful one is Ben Or's lower bound method for algorithms described as algebraic computation trees [BO83].

Here is a quick application from [AFR85] which uses the more precise bound in Theorem 10.4. Let the *sign pattern* of an $n \times n$ matrix A be the matrix S with $s_{ij} = \operatorname{sgn} a_{ij}$. We claim that there are $n \times n$ matrices S with only ± 1 entries such that every A with sign pattern S has rank at least cn, for a positive constant c.

On the one hand, there are 2^{n^2} possible S's. On the other hand, an A of rank at most r can be written as UV^T, where U and V are $n \times r$ matrices. We consider the $2nr$ entries of U and V as variables; then the signs of the entries of A are signs of quadratic polynomials in these variables. We have $m = n^2$ polynomials and thus, by Theorem 10.4, there are no more than $O(n^2/nr)^{2nr}$ possible sign patterns of a rank-r matrix A. For $r < cn$ and c small, this quantity is smaller than 2^{n^2}, and so some patterns force rank at least cn.

11 Literature

Textbooks and lecture notes for such a classical subject as algebraic geometry abound, of course, but not all of them are equally accessible to beginners.

The usual hands-on introduction, with emphasis on computational aspects, is Cox, Little, and O'Shea [CLO07]. Schenck's book [Sch03] is very clear, readable, and concise; another advantage is that it also treats many related concepts from algebra and topology. A very good set of lecture notes freely accessible on the web, including some of the more advanced concepts, such as sheaves and schemes, is Gathmann [Gat13].

For intersection theory, dealing with generalizations of Bézout's theorem and other counting questions for varieties, a remarkable little book is Katz [Kat06], and an older concise introduction is Fulton [Ful84].

For combinatorial, geometric, and computer science applications of polynomials, we can recommend, for example, Chen, Kayal, and Wigderson [CKW11]. Recent treatments of methods similar to the one used in the joints problem are Guth [Gut13] and Tao [Tao13].

Acknowledgment. I would like to thank Boris Bukh, Vincent Kusters, Zuzana Safernová, Adam Sheffer, and Noam Solomon for valuable comments, suggestions, and corrections to earlier versions of this document.

References

[AFR85] N. Alon, P. Frankl, and V. Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *Proc. 26th IEEE Symposium on Foundations of Computer Science*, pages 277–280, 1985.

[Arr06] E. Arrondo. Another elementary proof of the Nullstellensatz. *Amer. Math. Monthly*, 113(2):169–171, 2006.

[Bar13] S. Barone. Some quantitative results in real algebraic geometry. Preprint, arXiv:1307.8353, 2013.

[BO83] M. Ben-Or. Lower bounds for algebraic computation trees. In *Proc. 15th Annu. ACM Sympos. Theory Comput.*, pages 80–86, 1983.

[BPR96] S. Basu, R. Pollack, and M.-F. Roy. On the combinatorial and algebraic complexity of quantifier elimination. *J. ACM*, 43(6):1002–1045, 1996.

[BPR03] S. Basu, R. Pollack, and M.-F. Roy. *Algorithms in real algebraic geometry.* Algorithms and Computation in Mathematics 10. Springer, Berlin, 2003.

[BV07] S. Basu and N. N. Vorobjov. On the number of homotopy types of fibres of a definable map. *J. Lond. Math. Soc., II. Ser.*, 76(3):757–776, 2007.

[CKW11] Xi Chen, N. Kayal, and A. Wigderson. Partial derivatives in arithmetic complexity and beyond. *Found. Trends Theor. Comput. Sci.*, 6(1-2):1–138, 2011.

[CLO05] D. A. Cox, J. Little, and D. O'Shea. *Using algebraic geometry.* Springer, New York, 2005.

[CLO07] D. Cox, J. Little, and D. O'Shea. *Ideals, varieties, and algorithms.* Undergraduate Texts in Mathematics. Springer, New York, third edition, 2007.

[Ful84] W. Fulton. *Introduction to intersection theory in algebraic geometry*, volume 54 of *CBMS Regional Conference Series in Mathematics.* Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1984.

[Gat13] A. Gathmann. Algebraic geometry. Lecture Notes, TU Kaiserslautern, `http://www.mathematik.uni-kl.de/agag/mitglieder/professoren/gathmann/notes/alggeom/`, 2013.

[GK10] L. Guth and N. H. Katz. Algebraic methods in discrete analogs of the Kakeya problem. *Adv. Math.*, 225(5):2828–2839, 2010.

[Gut13] L. Guth. The polynomial method, 2013. Book in preparation.

[GV88] D. Yu. Grigor'ev and N. N. Vorobjov jun. Solving systems of polynomial inequalities in subexponential time. *J. Symb. Comput.*, 5(1-2):37–64, 1988.

[Har09] N. J. A. Harvey. Algebraic algorithms for matching and matroid problems. *SIAM J. Comput.*, 39(2):679–702, 2009.

[Hei83] J. Heintz. Definability and fast quantifier elimination in algebraically closed fields. *Theoret. Comput. Sci.*, 24(3):239–277, 1983. Corrigendum ibid. 39,1983: 2–3.

[Kat06] S. Katz. *Enumerative geometry and string theory*, volume 32 of *Student Mathematical Library*. American Mathematical Society, Providence, RI, 2006. IAS/Park City Mathematical Subseries.

[Kol88] J. Kollár. Sharp effective Nullstellensatz. *J. Amer. Math. Soc.*, 1(4):963–975, 1988.

[Mil64] J. W. Milnor. On the Betti numbers of real algebraic varieties. *Proc. Amer. Math. Soc.*, 15:275–280, 1964.

[MO02] C. Michaux and A. Ozturk. Quantifier elimination following Muchnik. Univ. de Mons-Hainaut Preprint Series (#10), 2002.

[Ole51] O. A. Oleinik. Estimates of the Betti numbers of real algebraic hypersurfaces (in Russian). *Mat. Sbornik (N. S.)*, 28(70):635–640, 1951.

[OP49] O. A. Oleinik and I. B. Petrovskiĭ. On the topology of of real algebraic surfaces (in Russian). *Izv. Akad. Nauk SSSR*, 13:389–402, 1949.

[Sch95] J. Schmid. On the affine Bézout inequality. *Manuscripta Math.*, 88(2):225–232, 1995.

[Sch03] H. Schenck. *Computational algebraic geometry*, volume 58 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2003.

[ST12] J. Solymosi and T. Tao. An incidence theorem in higher dimensions. *Discrete Comput. Geom.*, 48(2):255–280, 2012.

[Tao12] T. Tao. Spending symmetry. Book in preparation, draft available at `http://http://terrytao.wordpress.com/books/spending-symmetry/`, 2012.

[Tao13] T. Tao. Algebraic combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory. Preprint, arXiv:1310.6482, 2013.

[Tho65] R. Thom. On the homology of real algebraic varieties (in French). In S. S. Cairns, editor, *Differential and Combinatorial Topology.* Princeton Univ. Press, 1965.

[Wal79] M. Waldschmidt. *Transcendence methods*. Queen's University, 1979. Available at `http://www.math.jussieu.fr/~miw/articles/pdf/QueensPaper52.pdf`.

[Woo96] T. D. Wooley. A note on simultaneous congruences. *J. Number Theory*, 58(2):288–297, 1996.