The weak Bézout theorem

Jiří Matoušek

Rev. 9/IV/12 JM

We aim at a more or less self-contained proof of the following useful result:

Theorem 1 (Weak Bézout theorem over the reals) *Let $f, g \in \mathbb{R}[x, y]$ be bivariate polynomials. If f and g have no common factor (of degree at least 1), then the zero sets $Z(f)$ and $Z(g)$ intersect in at most $k\ell$ points, where $k := \deg f$ and $\ell := \deg g$.*

Remarks. The "usual" Bézout theorem asserts that polynomials f, g with no common factor have *exactly* $k\ell$ intersections of their zero sets, but this needs stronger assumptions: we need to work over an *algebraically closed field*, say $\mathbb{C}$; we need to count the intersections with appropriately defined *multiplicity*; and we need to consider zero sets in the *projective plane*, including points at infinity.

The proof of Theorem 1 given here works, with a minor modification, over any *infinite* field in place of $\mathbb{R}$. The theorem itself holds over finite fields as well, though.

We should also note that if $f, g \in \mathbb{R}[x, y]$ do have a common factor h, then $Z(h)$ may be finite or empty, say, so we cannot claim that the intersection of $Z(f)$ and $Z(g)$ is infinite.

The following proof is essentially extracted from the treatment in the book [D. Cox, J. Little and D. O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra* (3rd edition), Springer-Verlag, Heidelberg, 2007].

Resultants. Now let $\mathbb{K}$ be an arbitrary field. Before starting with the proof of the weak Bézout theorem, we introduce a useful tool, the resultant of two polynomials. Here, for a while, we will deal with *univariate* polynomials $f, g \in \mathbb{K}[x]$.

Writing $f(x) = \sum_{i=0}^{k} a_i x^i$ and $g(x) = \sum_{j=0}^{\ell} b_j x^j$, we define the *Sylvester matrix* of f and g. This is a $(k+\ell) \times (k+\ell)$ matrix made of the coefficients of f and g and 0's. The definition is perhaps best grasped from an example with specific values of k and ℓ, here $k = 5$ and $\ell = 3$:

$$\begin{pmatrix} a_0 & 0 & 0 & b_0 & 0 & 0 & 0 & 0 \\ a_1 & a_0 & 0 & b_1 & b_0 & 0 & 0 & 0 \\ a_2 & a_1 & a_0 & b_2 & b_1 & b_0 & 0 & 0 \\ a_3 & a_2 & a_1 & b_3 & b_2 & b_1 & b_0 & 0 \\ a_4 & a_3 & a_2 & 0 & b_3 & b_2 & b_1 & b_0 \\ a_5 & a_4 & a_3 & 0 & 0 & b_3 & b_2 & b_1 \\ 0 & a_5 & a_4 & 0 & 0 & 0 & b_3 & b_2 \\ 0 & 0 & a_5 & 0 & 0 & 0 & 0 & b_3 \end{pmatrix}.$$

The *resultant* of f and g, w.r.t. the variable x, is denoted by $\mathrm{Res}(f, g, x)$ and defined as the determinant of the Sylvester matrix of f and g. Thus, $\mathrm{Res}(f, g, x)$ is an element of the field $\mathbb{K}$, and if we regard the coefficients $a_0, \ldots, a_k$ and $b_0, \ldots, b_\ell$ as variables, then $\mathrm{Res}(f, g, x)$ is a polynomial in these variables.

Lemma 2 *Two polynomials $f, g \in \mathbb{K}[x]$ have a common factor $h \in \mathbb{K}[x]$ (of degree at least 1) if and only if* $\mathrm{Res}(f, g, x) = 0$.

Proof. We will prove that both of the statements in the lemma are equivalent to the following statement (*):

> *There exist polynomials $A, B \in \mathbb{K}[x]$, not both zero, such that $\deg A \le \ell - 1$, $\deg B \le k - 1$, and $Af + Bg = 0$ (where $k = \deg f$ and $\ell = \deg g$).*[1]

The equivalence of (*) with $\mathrm{Res}(f, g, x) = 0$ is a simple linear algebra. Let us write $A = A(x) = \sum_{i=0}^{\ell-1} \alpha_i x^i$, $B = B(x) = \sum_{j=0}^{k-1} \beta_j x^j$, and let us regard the coefficients $\alpha_0, \ldots, \alpha_{\ell-1}$ and $\beta_0, \ldots, \beta_{k-1}$ as unknowns. Then the condition $Af + Bg = 0$ translates to a system of $k + \ell$ homogeneous linear equations for these $k + \ell$ unknowns, and the matrix of this system is precisely the Sylvester matrix of f and g.

As is well known, a homogeneous linear system $M\mathbf{x} = \mathbf{0}$ with a square matrix M has a nonzero solution iff $\det M = 0$. In our case, a nonzero solution is equivalent to the existence of A, B as in the statement (*).

Next, let us assume that f and g have a common factor h, $\deg h \ge 1$, and write $f = hf_1$, $g = hg_1$. Then $A := g_1$ and $B := -f_1$ satisfy $Af + Bg = g_1 h f_1 - f_1 h g_1 = 0$, and so (*) holds.

Conversely, suppose that f, g have no nontrivial common factor. Then, as is well known, there are polynomials $u, v \in \mathbb{K}[x]$ with $uf + vg = 1$.[2] Let us

[1] The equality $Af + Bg = 0$ is meant as equality of *polynomials*, i.e., elements of $\mathbb{K}[x]$. Note that for finite field $\mathbb{K}$, a nonzero polynomial may represent the zero function—for example, this is the case for the polynomial $x^2 - x$ over the two-element field.

[2] For a quick proof, we consider a polynomial h that has the smallest possible degree among all nonzero polynomials of the form $uf + vg$, and we want to check that h divides both f and g (and consequently, h is a greatest common divisor of f and g). We can write $f = qh + r$ for suitable polynomials q and r, with $\deg r < \deg h$ (this is division with remainder). If $r = 0$, then f is a multiple of h as needed, and otherwise, we express $r = f - qh = (1 - qu)f - (qv)g$, and we get a contradiction to the choice of h.

suppose that $A, B \in \mathbb{K}[x]$ satisfy $Af + Bg = 0$ as in (*); we want to check that at least one of them has too large degree.

Using $Af = -Bg$ we compute $A = A \cdot 1 = Auf + Avg = -uBg + Avg = (Au - Bv)g$. Thus, if $A \neq 0$, then $\deg A \geq \deg g = \ell$. Similarly we find that if $B \neq 0$, then $\deg B \geq \deg f = k$. So the statement (*) cannot hold. The lemma is proved. □

Remark. If $\mathbb{K}$ is an algebraically closed field, such as $\mathbb{C}$, then the polynomials f and g factor as $f(x) = a_k \prod_{i=1}^{k}(x - \rho_i)$, $g(x) = b_\ell \prod_{j=1}^{\ell}(x - \sigma_j)$, where the ρ_i are the roots of f and the σ_j the roots of g. Then it can be shown that

$$\mathrm{Res}(f, g, x) = a_k b_\ell \prod_{i=1}^{k} \prod_{j=1}^{\ell} (\rho_i - \sigma_j).$$

This formula, which we won't prove, makes it clear that the resultant vanishes iff f and g have a common root.

Back to bivariate polynomials. Now we again consider two polynomials $f, g \in \mathbb{R}[x, y]$. In order to be able to use resultants, we will regard f and g as polynomials in x with coefficients in $\mathbb{R}[y]$; that is, $f = \sum_{i=0}^{k} a_i(y) x^i$, where each $a_i(y)$ is a polynomial in y.

Then the entries of the Sylvester matrix of f and g are polynomials in y, and so $\mathrm{Res}(f, g, x)$, which is the determinant of a matrix of polynomials, is still well-defined.

Lemma 3 *For $f, g \in \mathbb{R}[x, y]$ with $\deg f = k$, $\deg g = \ell$, we have $\deg \mathrm{Res}(f, g, x) \leq k\ell$.*

We leave the proof as an exercise (not entirely trivial). Hint: consider the terms in the expansion of the determinant (as a sum over all permutations), and observe that, with $f = \sum_{i=0}^{k} a_i(y) x^i$ as above, $\deg a_i(y) \leq k - i$.

Proof of Theorem 1. Let us suppose that $Z(f) \cap Z(g)$ has at least $k\ell + 1$ points, and let us fix some points $p_1, \ldots, p_{k\ell+1} \in Z(f) \cap Z(g)$.

First we want to rotate the coordinate system so that all the p_i have distinct y-coordinates. This is possible since there are only finitely many lines containing two or more of the p_i, and if the x-axis is not parallel to any of these lines, then the y-coordinates are all distinct.

Algebraically speaking, rotating the coordinate system by some suitable angle α means replacing the old coordinates (x, y) by new coordinates (x^*, y^*), where $x = ax^* + by^*$, $y = cx^* + dy^*$, with $a = c = \cos\alpha$, $b = -d = \sin\alpha$. By this substitution, we obtain new polynomials $f^*(x^*, y^*) = f(ax^* + by^*, cx^* + dy^*)$, $g^*(x^*, y^*) = g(ax^* + by^*, cx^* + dy^*)$. It is easily checked that $\deg f^* = \deg f$, $\deg g^* = \deg g$, and that f and g have common factor iff f^* and g^* have one (for this, we need that the substitution is invertible, i.e., x^* and y^* can be expressed as linear functions of x and y).

Thus, from now on, we assume that the y-coordinates $y_1, \ldots, y_{k\ell+1}$ of $p_1, \ldots, p_{k\ell+1}$ are all distinct, and for simpler notation, we keep calling our polynomials f and g.

Now for each y_i, $f_i(x) := f(x, y_i)$ and $g_i(x) := g(x, y_i)$ are univariate polynomials, and they have a common root, namely, the x-coordinate of p_i. Having a common root implies having a common factor, and hence $\mathrm{Res}(f_i, g_i, x) = 0$.

Since $\mathrm{Res}(f_i, g_i, x)$ is the value of the polynomial $\mathrm{Res}(f, g, x)$ at y_i, we conclude that $\mathrm{Res}(f, g, x)$ has at least $k\ell + 1$ distinct roots. By Lemma 3 we know that $\deg \mathrm{Res}(f, g, x) \leq k\ell$, and so $\mathrm{Res}(f, g, x)$ is the zero polynomial.

Now we would like to conclude that since the resultant is zero, f and g have a common factor. But there is a catch: Lemma 2 assumes that the coefficients of the considered polynomials belong to a *field* $\mathbb{K}$, but in our case, the coefficients are from $\mathbb{R}[y]$, which most certainly is not a field!

A way around this is extending $\mathbb{R}[y]$ into a field. Namely, one can imitate the usual construction of the field of rational numbers from the ring of integers. In the case of $\mathbb{R}[y]$ we arrive at the field $\mathbb{R}(y)$, consisting of all *rational functions* of the form $p(y)/q(y)$, where $q(y)$ is nonzero. (More precisely, the elements of $\mathbb{R}(y)$ are equivalence classes of rational functions, similarly as, e.g., $\frac{3}{4}$ represents the same fraction as $\frac{6}{8}$.)

So for the purpose of using Lemma 2, we regard f, g as polynomials in x with coefficients in the field $\mathbb{R}(y)$. Then $\mathrm{Res}(f, g, x)$, being the zero polynomial, is also the zero element of $\mathbb{R}(y)$, so Lemma 2 allows us to conclude that f and g have a common factor $h \in \mathbb{R}(y)[x]$. That is, we can write $f = hf_1$ and $g = hg_1$, where the coefficients of h, f_1, g_1 are rational functions in y.

Let $d = d(y)$ be a common denominator of all the coefficients of h ,f_1, and g_1; thus, we can write

$$h(x, y) = \frac{\tilde{h}(x, y)}{d(y)}, \quad f_1(x, y) = \frac{\tilde{f}_1(x, y)}{d(y)}, \quad g_1(x, y) = \frac{\tilde{g}_1(x, y)}{d(y)},$$

where $\tilde{h}, \tilde{f}_1, \tilde{g}_1$ are polynomials with coefficients in $\mathbb{R}$ (no rational functions anymore).

We let $\tilde{h}_1$ be an irreducible factor of $\tilde{h}$ of degree at least 1 in x (there must be such a factor since h contains a nonzero power of x). From the equality $f = hf_1 = \tilde{h}\tilde{f}_1/d^2$ we see that the irreducible factor $\tilde{h}_1$ has to divide the product fd^2. It cannot divide d^2, since d contains no power of x, and hence $\tilde{h}_1$ has to divide f. The same argument shows that $\tilde{h}_1$ divides g as well, and hence we can finally conclude that f and g have a common factor.

In the last argument, we have used the *unique factorization property* of the polynomial ring $\mathbb{R}[x, y]$, stating that every polynomial in $\mathbb{R}[x, y]$ can be factorized into irreducible factors, in a way that is unique up to reordering the factors and multiplying them by nonzero numbers (the same holds for $\mathbb{K}[x, y]$ with any field $\mathbb{K}$). This property is well known; a full proof would take perhaps another page, but here we skip it. □