3. Resultants

In this chapter we present resultants, another method of elimination theory. Resultants are historically older than Gröbner bases. They are somehow easier to compute, but on the other hand they do not generate as much information as a Gröbner basis.

Theorem 3.1. (B.L.van der Waerden, "Algebra, vol.I", p.102)
Let $a(x), b(x)$ be two non-constant polynomials in $K[x]$, K a field. Then a and b have a non-constant common factor (i.e. a common root over the algebraic closure of K) if and only if there are polynomials $p(x), q(x) \in K[x]$, not both equal to 0, with $\deg(p) < \deg(b), \deg(q) < \deg(a)$, *such that*

$$p(x)a(x) + q(x)b(x) = 0 \ . \qquad (*)$$

Proof: If a and b have the non-constant common factor c, then obviously we can write

$$(b/c) \cdot a - (a/c) \cdot b = 0 \ .$$

On the other hand, assume $(*)$. So we have

$$p(x)a(x) = -q(x)b(x) \ . \qquad (**)$$

We factor the left and right hand sides of $(**)$ into irreducible factors. All the irreducible factors of $a(x)$ must divide the right hand side at least as often as they divide $a(x)$. Yet they cannot divide $q(x)$ as often as they do $a(x)$ because of the degree restriction. Hence at least one irreducible factor of $a(x)$ occurs also in $b(x)$. □

How can we decide the existence of such polynomials p and q as in the previous theorem?

Let $m = \deg(a), n = \deg(b)$ and write

$$a(x) = \sum_{i=0}^{m} a_i x^i, \qquad b(x) = \sum_{i=0}^{n} b_i x^i \ .$$

Ansatz:

$$p(x) = \sum_{i=0}^{n-1} p_i x^i, \qquad q(x) = \sum_{i=0}^{m-1} q_i x^i \ .$$

Then

$$
\begin{array}{l}
p \cdot a + q \cdot b = 0 \\
\Longleftrightarrow \\
\mathrm{coeff}(p \cdot a, x^i) + \mathrm{coeff}(q \cdot b, x^i) = 0 \quad \forall i \\
\Longleftrightarrow \\
\qquad p_{n-1}a_m + q_{m-1}b_n = 0 \\
\qquad\qquad\quad \vdots \\
p_0 a_1 + p_1 a_0 + q_0 b_1 + q_1 b_0 = 0 \\
\qquad\quad p_0 a_0 + q_0 b_0 = 0 \\
\Longleftrightarrow \\
(p_{n-1}, \ldots, p_0, q_{m-1}, \ldots, q_0) \cdot \begin{pmatrix} a_m & \cdots & a_0 & & \\ & \ddots & & \ddots & \\ & & a_m & \cdots & a_0 \\ b_n & \cdots & b_0 & & \\ & \ddots & & \ddots & \\ & & b_n & \cdots & b_0 \end{pmatrix} = (0, \ldots, 0) \ .
\end{array}
$$

So there is a non-trivial solution for p and q if and only if the matrix in this equation has determinant 0.

Definition 3.2. Let

$$a(x) = \sum_{i=0}^{m} a_i x^i, \qquad b(x) = \sum_{i=0}^{n} b_i x^i$$

be non-constant polynomials in $I[x]$ (I an integral domain) of degree m and n, respectively. Let $\mathrm{Syl}_x(a, b)$ be the **Sylvester matrix** of a and b, i.e.

$$
\mathrm{Syl}_x(a,b) \ = \ \begin{pmatrix}
a_m & a_{m-1} & \cdots & \cdots & a_1 & a_0 & 0 & \cdots & \cdots & 0 \\
0 & a_m & a_{m-1} & \cdots & \cdots & a_1 & a_0 & 0 & \cdots & 0 \\
 & & & \vdots & & & & & & \\
0 & \cdots & \cdots & 0 & a_m & a_{m-1} & \cdots & \cdots & a_1 & a_0 \\
- & - & - & - & - & - & - & - & - & - \\
b_n & b_{n-1} & \cdots & \cdots & b_1 & b_0 & 0 & \cdots & \cdots & 0 \\
0 & b_n & b_{n-1} & \cdots & \cdots & b_1 & b_0 & 0 & \cdots & 0 \\
 & & & \vdots & & & & & & \\
0 & \cdots & \cdots & 0 & b_n & b_{n-1} & \cdots & \cdots & b_1 & b_0
\end{pmatrix} \ .
$$

The lines of $\mathrm{Syl}_x(a, b)$ consist of the coefficients of the polynomials $x^{n-1}a(x), \ldots, xa(x), a(x)$ and $x^{m-1}b(x), \ldots, xb(x), b(x)$, i.e. there are n lines of coefficients of a and m lines of coefficients of b. The **resultant** of a and b is the determinant of $\mathrm{Syl}_x(a, b)$; i.e.

$$\mathrm{res}_x(a, b) := \det(\mathrm{Syl}_x(a, b) \ . \qquad \square$$

The *resultant* $\mathrm{res}_x(f, g)$ of two univariate polynomials $f(x), g(x)$ over an integral domain I is the determinant of the *Sylvester matrix* of f and g, consisting of shifted lines of

coefficients of f and g. $\mathrm{res}_x(f,g)$ is a constant in I. For $m = \deg(f), n = \deg(g)$, we have $\mathrm{res}_x(f,g) = (-1)^{mn}\mathrm{res}_x(g,f)$, i.e. the resultant is symmetric up to sign. If $a_1, \ldots, a_m$ are the roots of f, and $b_1, \ldots, b_n$ are the roots of g in their common splitting field, then

$$\mathrm{res}_x(f,g) = \mathrm{lc}(f)^n \mathrm{lc}(g)^m \prod_{i=1}^{m}\prod_{j=1}^{n}(a_i - b_j).$$

The resultant has the important property that, for non–zero polynomials f and g, $\mathrm{res}_x(f,g) = 0$ if and only if f and g have a common root, and in fact, f and g have a non-constant common divisor in $K[x]$, where K is the quotient field of I. If f and g have positive degrees, then there exist polynomials $a(x), b(x)$ over I such that $af + bg = \mathrm{res}_x(f,g)$. The *discriminant* of $f(x)$ is

$$\mathrm{discr}_x(f) = (-1)^{m(m-1)/2}\mathrm{lc}(f)^{2(m-1)} \prod_{i\neq j}(a_i - a_j).$$

We have the relation $\mathrm{res}_x(f,f') = (-1)^{m(m-1)/2}\mathrm{lc}(f)\mathrm{discr}_x(f)$, where f' is the derivative of f.

The resultant of f and g can be written as a linear combination of f and g. This is proven in [CLO98] [1] for polynomials over a field. But the proof can be extended to polynomials over integral domains.

Theorem 3.3. *Given $a, b \in I[x]$ of positive degree, where I is an integral domain. Then*

$$\mathrm{res}_x(a,b) = p \cdot a + q \cdot b$$

for some $p(x), q(x) \in I[x]$. So $\mathrm{res}_x(a,b)$ is in the ideal generated by a and b in $I[x]$.

Proof: Let $m = \deg(a), n = \deg(b)$. The definition of resultant was based on the equation $p \cdot a + q \cdot b = 0$. In this proof, we will apply the same method to the equation

$$\tilde{p} \cdot a + \tilde{q} b = 1\ . \qquad (*)$$

The theorem is trivially true if $\mathrm{res}_x(a,b) = 0$; simply choose $p = q = 0$.
So we may assume $\mathrm{res}_x(a,b) \neq 0$. Now let

$$a = \sum_{i=0}^{m} a_i x^i\ , \quad b = \sum_{i=0}^{n} b_i x^i\ , \quad \tilde{p} = \sum_{i=0}^{n-1} \tilde{p}_i x^i\ , \quad \tilde{q} = \sum_{i=0}^{m-1} \tilde{q}_i x^i\ ,$$

where the coefficients of $\tilde{p}$ and $\tilde{q}$ are unknowns in I. If we substitute these formulas into $(*)$ and compare coefficients of powers of x, then we get a system of linear equations with coefficients a_i, b_i and with unknowns c_i, d_i. These equations are the same as in the derivation of the Sylvester matrix, except for 1 on the right hand side for the coefficient of x^0. So we have

$$\mathrm{Syl}_x(a,b)^T \cdot \begin{pmatrix} \tilde{p}_{n-1} \\ \vdots \\ \tilde{p}_0 \\ \tilde{q}_{m-1} \\ \vdots \\ \tilde{q}_0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

[1] Cox,Little,O'Shea, "Ideals, Varieties, and Algorithms", 2nd ed., p.152

Because the resultant is non-zero, this system has a unique solution. So we can use Cramer's rule to give a formula for the components of the solution. For example,

$$\tilde{p}_{n-1} = \frac{\det(M_1)}{\mathrm{res}_x(a,b)} \ ,$$

where the matrix M_1 is the result of replacing the first column in $\mathrm{Syl}_x(a,b)^T$ by the right hand side $(0\ldots0\,1)^T$. The numerator is an element in the integral domain I. We get analogous expressions for the other unknowns. So we can write

$$\tilde{p} = \frac{1}{\mathrm{res}_x(a,b)} \cdot p \ , \quad \tilde{q} = \frac{1}{\mathrm{res}_x(a,b)} \cdot q \ ,$$

where p and q are in $I[x]$. Since $\tilde{p}$ and $\tilde{q}$ safisfy $\tilde{p} \cdot a + \tilde{q} \cdot b = 1$, we can multiply through by $\mathrm{res}_x(a,b)$ to obtain

$$p \cdot a + q \cdot b = \mathrm{res}_x(a,b) \, .$$

So $\mathrm{res}_x(a,b)$ is in the ideal generated by a and b in $I[x]$. □

For actually computing resultants, e.g. for polynomials in $\mathbb{Z}[x_1,\ldots,x_n]$, one uses a modular approach similar to the one for gcd computation. So some of the variables are evaluated at several evaluation points and the finial result is then interpolated. We state a crucial Lemma needed for this process.

Lemma 3.4. (Lemma 4.3.1 in Winkler, "Computer Algebra")
Let I, J be integral domains, ϕ a homomorphism from I into J. The homomorphism from $I[x]$ into $J[x]$ induced by ϕ will also be denoted ϕ, i.e. $\phi(\sum_{i=0}^m c_i x^i) = \sum_{i=0}^m \phi(c_i) x^i$. Let $a(x), b(x)$ be polynomials in $I[x]$. If $\deg(\phi(a)) = \deg(a)$ and $\deg(\phi(b)) = \deg(b) - k$, then $\phi(\mathrm{res}_x(a,b)) = \phi(\mathrm{lc}(a))^k \mathrm{res}_x(\phi(a), \phi(b))$.

Proof: Let M be the Sylvester matrix of a and b, M^* the Sylvester matrix of $a^* = \phi(a)$ and $b^* = \phi(b)$. If $k = 0$, then clearly $\phi(\mathrm{res}_x(a,b)) = \mathrm{res}_x(a^*, b^*)$.
If $k > 0$ then M^* can be obtained from $\phi(M)$ by deleting its first k rows and columns. Since the first k columns of $\phi(M)$ contain $\phi(\mathrm{lc}(a))$ on the diagonal and are zero below the diagonal, $\phi(\mathrm{res}_x(a,b)) = \phi(\det(M)) = \det(\phi(M)) = \phi(\mathrm{lc}(a))^k \mathrm{res}_x(a^*, b^*)$. □

Solving systems of algebraic equations by resultants

Theorem 3.5. (Theorem 4.3.3 in Winkler, "Computer Algebra")
Let K be an algebraically closed field, let

$$\begin{aligned} a(x_1,\ldots,x_r) &= \textstyle\sum_{i=0}^{m} a_i(x_1,\ldots,x_{r-1})x_r^i, \\ b(x_1,\ldots,x_r) &= \textstyle\sum_{i=0}^{n} b_i(x_1,\ldots,x_{r-1})x_r^i \end{aligned}$$

be elements of $K[x_1,\ldots,x_r]$ of positive degrees m and n in x_r, and let $c(x_1,\ldots,x_{r-1}) = \mathrm{res}_{x_r}(a,b)$. If $(\alpha_1,\ldots,\alpha_r) \in K^r$ is a common root of a and b, then $c(\alpha_1,\ldots,\alpha_{r-1}) = 0$. Conversely, if $c(\alpha_1,\ldots,\alpha_{r-1}) = 0$, then one of the following holds:
(a) *$a_m(\alpha_1,\ldots,\alpha_{r-1}) = b_n(\alpha_1,\ldots,\alpha_{r-1}) = 0$,*
(b) *for some $\alpha_r \in K$, $(\alpha_1,\ldots,\alpha_r)$ is a common root of a and b.*

Proof: By Theorem 3.3 we have $c = ua + vb$, for some $u, v \in K[x_1,\ldots,x_r]$. If $(\alpha_1,\ldots,\alpha_r)$ is a common root of a and b, then the evaluation of both sides of this equation immediately yields $c(\alpha_1,\ldots,\alpha_{r-1}) = 0$.

Now assume $c(\alpha_1,\ldots,\alpha_{r-1}) = 0$. Suppose $a_m(\alpha_1,\ldots,\alpha_{r-1}) \neq 0$, so we are not in case (a). Let ϕ be the evaluation homomorphism $x_1 = \alpha_1, \ldots, x_{r-1} = \alpha_{r-1}$. Let $k = \deg(b) - \deg(\phi(b))$. By Lemma 3.4. we have $0 = c(\alpha_1,\ldots,\alpha_{r-1}) = \phi(c) = \phi(\mathrm{res}_{x_r}(a,b)) = \phi(a_m)^k \mathrm{res}_{x_r}(\phi(a),\phi(b))$. Since $\phi(a_m) \neq 0$, we have $\mathrm{res}_{x_r}(\phi(a),\phi(b)) = 0$. Since the leading term in $\phi(a)$ is non–zero, $\phi(a)$ and $\phi(b)$ must have a common non–constant factor, say $d(x_r)$ (see (van der Waerden 1970), Sec. 5.8). Let α_r be a root of d in K. Then $(\alpha_1,\ldots,\alpha_r)$ is a common root of a and b. Analogously we can show that (b) holds if $b_n(\alpha_1,\ldots,\alpha_{r-1}) \neq 0$. □

This theorem suggests a method for determining the solutions of a system of algebraic, i.e. polynomial, equations over an algebraically closed field. Suppose, for example, that a system of three algebraic equations is given as

$$a_1(x,y,z) = a_2(x,y,z) = a_3(x,y,z) = 0.$$

Let, e.g.,

$$\begin{aligned} b(x) &= \mathrm{res}_z(\mathrm{res}_y(a_1,a_2),\mathrm{res}_y(a_1,a_3)), \\ c(y) &= \mathrm{res}_z(\mathrm{res}_x(a_1,a_2),\mathrm{res}_x(a_1,a_3)), \\ d(z) &= \mathrm{res}_y(\mathrm{res}_x(a_1,a_2),\mathrm{res}_x(a_1,a_3)). \end{aligned}$$

In fact, we might compute these resultants in any other order. By Theorem 3.5, all the roots $(\alpha_1,\alpha_2,\alpha_3)$ of the system satisfy $b(\alpha_1) = c(\alpha_2) = d(\alpha_3) = 0$. So if there are finitely many solutions, we can check for all of the candidates whether they actually solve the system.

Unfortunately, there might be solutions of b, c, or d, which cannot be extended to solutions of the original system, as we can see from the following example.

Example 3.6. Consider the system of algebraic equations

$$\begin{aligned} a_1(x,y,z) &= 2xy + yz - 3z^2 = 0, \\ a_2(x,y,z) &= x^2 - xy + y^2 - 1 = 0, \\ a_3(x,y,z) &= yz + x^2 - 2z^2 = 0. \end{aligned}$$

We compute

$$\begin{aligned} b(x) &= \mathrm{res}_z(\mathrm{res}_y(a_1,a_3),\mathrm{res}_y(a_2,a_3)) \\ &= x^6(x-1)(x+1)(127x^4-167x^2+4), \\ c(y) &= \mathrm{res}_z(\mathrm{res}_x(a_1,a_3),\mathrm{res}_x(a_2,a_3)) \\ &= (y-1)^3(y+1)^3(3y^2-1)(127y^4-216y^2+81)\cdot(457y^4-486y^2+81), \\ d(z) &= \mathrm{res}_y(\mathrm{res}_x(a_1,a_2),\mathrm{res}_x(a_1,a_3)) \\ &= 5184z^{10}(z-1)(z+1)(127z^4-91z^2+16). \end{aligned}$$

All the solutions of the system, e.g. $(1,1,1)$, have coordinates which are roots of b,c,d. But there is no solution of the system having y–coordinate $1/\sqrt{3}$. So not every root of these resultants can be extended to a solution of the system.

The Göbner basis of the ideal generated by a_1,a_2,a_3 w.r.t. lexicographic ordering with $z>x>y$ contains the univariate polynomial

$$g_1(y) = (y-1)(y+1)(127y^4-216y^2+81)\ .$$

So no extraneous factors are generated. All solutions of $g_1(y)$ can be extended to solutions of the whole system. □

Example 3.7. According to Theorem 3.5(a), a partial solution $(\alpha_1,\ldots,\alpha_{r-1})$ might not be extendable to a full common solution of a and b, if it is a common root of the leading coefficients of a and b. But in certain cases it might still be extendable. As an example consider

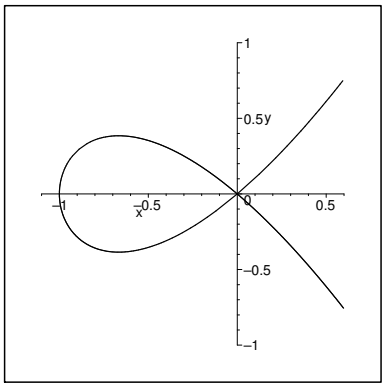
$$\begin{aligned} a(x_1,x_2) &= -x_1x_2^2+x_1^2x_2-4x_2+x_1\ , \\ b(x_1,x_2) &= 2x_1x_2^2+x_1^2x_2-3x_1x_2-4x_2+x_1\ . \end{aligned}$$

The resultant w.r.t. x_2 is

$$c(x_1) = \mathrm{res}_{x_2}(a,b) = 9x_1^3(x_1-2)(x_1+2)\ .$$

The leading coefficients of a and b w.r.t. x_2 are $-x_1$ and $2x_1$, respectively. They both vanish at 0, but still $(0,0)$ is a common solution of a and b. □

Example 3.8. Let $\mathcal{C}$ be the affine curve (node) defined by $f(x,y)=x^3+x^2-y^2=0$.



$\mathcal{C}$ has a double point at the origin $O=(0,0)$. Intersecting $\mathcal{C}$ by the line $\mathcal{L}$ defined by $l(x,y)=y-tx=0$,

$$\begin{aligned} \mathrm{res}_y(f,l) &= x^2\cdot(x-t^2+1)\ , \\ \mathrm{res}_x(f,l) &= y^2\cdot(y-t^3+t)\ , \end{aligned}$$

we get the additional intersection point (t^2-1, t^3-t). So

$$x(t) = t^2 - 1, \qquad y(t) = t^3 - t$$

is a parametrization of $\mathcal{C}$. □