6.966 Algebra and Computation	September 21, 1998
Lecture 4	
Lecturer: Madhu Sudan	*Scribe: Leonid Reyzin*

Resultant

First we recall the definition of a resultant from the previous lecture. Let $a, b \in F[x]$ be polynomials of degrees m, n, respectively. We can ask a question of whether $(a, b) = 1$ or, equivalently, whether there exist two polynomials u, v of degrees less than n, m, respectively, such that $au + bv = 1$ (here and below we use (a, b) to denote a greatest common divisor of a and b). This question can be considered as a system of $m+n$ linear equations in $m + n$ variables, where the variables are the n coefficients of u and m coefficients of v, and the coefficients for the equations come from the coefficients of a and b. We define $\mathrm{Res}(a, b)$ to be the determinant of the matrix for this system.

More precisely, let $a = \sum_{i=0}^{m} a_i x^i$ and $b = \sum_{i=0}^{n} b_i x^i$ and $M_{a,b}$ be the $(n + m) \times (n + m)$ matrix given by

$$(M_{a,b})_{ij} = \begin{cases} a_{i-j} & \text{if } j \le n \text{ and } 0 \le i - j \le m. \\ b_{i-(j-n)} & \text{if } j > n \text{ and } 0 \le i - (j - n) \le n. \\ 0 & \text{otherwise.} \end{cases}$$

Let w be the $m + n$-dimensional vector of unknowns. I.e., the ith coordinate of w is set to u_{i-1} if $i \le n$ and v_{i-n-1} if $i > n$, where the u_j's and v_j's are the coefficients of u and v. Then the linear system to be solved is given by $M_{a,b}w = (100\ldots0)^T$. The resultant of a and b, $\mathrm{Res}(a, b)$, is defined to be the determinant of the matrix $M_{a,b}$.

Theorem 1 $\mathrm{Res}(a, b) \neq 0$ *if and only if* $(a, b) = 1$.

Proof The forward direction is easy: if $\mathrm{Res}(a, b) \neq 0$, then the linear system $M_{a,b}w = (100\ldots0)^T$ has a solution (by Cramer's rule), so $\exists u, v$ such that $au + bv = 1$, so $(a, b) = 1$.

To show the other direction, assume $(a, b) = 1$. Then $\exists u, v \in F[x]$ of degrees less than n, m, respecitvely, such that $au + bv = 1$. We also claim that for all i, $0 \le i < m + n$, $\exists u^{(i)}, v^{(i)} \in F[x]$ of degrees less than n, m, respectively, such that $au^{(i)} + bv^{(i)} = x^i$. This is because we can write, using the division algorithm, $ux^i = qb + r$ for $\deg(r) < \deg(b)$ and set $u^{(i)} = r = ux^i - qb$ and $v^{(i)} = vx^i + qa$. It is clear that $au^{(i)} + bv^{(i)} = aux^i - qab + bvx^i + qab = x^i(au + bv) = x^i$. It is also clear that $\deg(u^{(i)}) < \deg(b) = n$. To see that $\deg(v^{(i)}) < m$, note that $au^{(i)}$ has degree at most $m + n - 1$, therefore, $bv^{(i)}$ has to have degree at most $m+n-1$ (otherwise the equation $au^{(i)} + bv^{(i)} = x^i$ could not possibly hold, because $i \le m+n-1$), so $\deg(v^{(i)}) \le m + n - 1 - \deg(b) = m - 1$. Thus, the claim holds. The claim implies that the system of linear equations in the coefficients of u and v has a solution for any right-hand side (not just for $(1, 0, 0, \ldots, 0)^T$, so the matrix of the system has to have full rank, so its determinant $\mathrm{Res}(a, b) \neq 0$. ■

Another way to view the resultant is as follows. Consider the case when a and b are not relatively prime: $\exists g \in F[x]$ such that $\deg(g) > 0$ and $(a, b) = g$. Then let $p = b/g$ and $q = -a/g$. Then $pa + qb = 0$. Thus, if a and b are not relatively prime, there exist $p, q \in F[x]$, $(p, q) \neq (0, 0)$ of degrees less than n, m, respectively, such that $pa + qb = 0$. Conversely, if such p and q exist, then a and b are not relatively prime (because if $pa = -qb$, then $b|pa$; but $\deg(p) < \deg(b)$ and $p \neq 0$, so a and b have to have a non-trivial common factor because $F[x]$ is a unique factorization domain). In this case we get the linear system $M_{a,b}w = \vec{0}$, where w now represents the coefficients of p and q rather than u and v. The primary difference from the previous case is that now the right-hand side is 0. a and b are not relatively prime if and only if a non-trivial solution exists, which is equivalent to saying that the matrix doesn't have full rank, or $\mathrm{Res}(a, b) = 0$.

All the discussions about the resultant and the GCD so far have assumed that F is a field. What happens if we relax this assumption? The resultant can be still be defined in the same way for polynomials over arbitrary commutative rings; our proofs, however, relied on for the properties of the field and may not hold for arbitrary rings. We will now consider the case of general unique factorization domains I. Since I is an integral domain, it is possible to consider the quotient field Q of I and use properties of $Q[x]$ to infer properties of $I[x]$.

Lemma 2 *(A variant of Gauss's lemma.) Two polynomials $a, b \in I[x]$ share a common factor of degree greater than 0 if and only if they share a common factor in $Q[x]$ of degree greater than 0.*

The discussions below show how to prove this lemma. (See the appendix of the notes for Lecture 2 for more details.)

Consider now the case of $I = \mathbb{Z}$, the integers. Then Q is the rationals. Two polynomials a, b with integer coefficients are relatively prime in $Q[x]$ if and only if $\exists u', v' \in Q[x]$ such that $av' + bu' = 1$ and $\deg(u') < \deg(a)$ and $\deg(v') < \deg(b)$. Equivalently, we can multiply through by the least common denominator N of all the coefficients of u' and v', to get that a, b don't share a factor of degree greater than 0 in $\mathbb{Z}[x]$ if and only if $\exists u, v \in \mathbb{Z}[x], N \in \mathbb{Z}$ such that $av + bu = N$ and $\deg(u) < \deg(a)$ and $\deg(v) < \deg(b)$.

How large can this N be? Let's assume that a, b are of degrees m, n, respectively, and their coefficients are at most l-bit integers. Remember that coefficients of u' and v' are given by Cramer's rule, which has $\mathrm{Res}(a, b)$ in the denominator. Thus, $0 < N \leq \mathrm{Res}(a, b)$. Remember also that $\mathrm{Res}(a, b)$ is the determinant of an $(m+n) \times (m+n)$ matrix with entries that are at most l-bits each, so N has at most $l(m+n)\log((m+n)!) = (m+n)^{O(1)}l$ bits. This leads to the following proposition.

Proposition 3 *Let $a, b \in \mathbb{Z}[x]$ be of degree m, n respectively, with coefficients of a, b being l-bit integers. Then there exist polynomials $u, v \in \mathbb{Z}[x]$ and $N \in \mathbb{Z} - \{0\}$ such that $ua + vb = N$ iff a and b have no common factor of positive degree in x. Furthermore if N as claimed above exists then there exists one that is at most $(m+n)!2^{(m+n)l}$.*

We can go through the same exercise for bivariate polynomials $F[x, y]$ over a field F. We just need to consider $F[x, y]$ as $F[x][y]$. Its elements are polynomials in y over an integral domain $F[x]$. We can consider the quotient field $F(x)$ of $F[x]$ (note the significant difference between parentheses and brackets in this notation). We get that $a(y), b(y)$ with coefficients in $F[x]$ are relatively prime in $F(x)[y]$ if and only if $\exists u'(y), v'(y) \in F(x)[y]$ such that $av' + bu' = 1$ and $\deg(u') < \deg(a)$ and $\deg(v') < \deg(b)$. Again, multiplying by the least common denominator $p(x)$ of the coefficients of u' and v', we get that $a(x, y)$ and $b(x, y)$ don't share a factor in $F[x, y]$ of degree in y greater than 0 if and only if $\exists u, v \in F[x, y]$ and $\exists p \in F[x]$ such that $au + bv = p$.

Again we might ask how large p is. Let's assume that the degrees of a and b in x are no more than d, and that their degrees in y are m and n, respectively. By the same reasoning as before, $\deg(p(x)) \leq (m+n)d$.

Proposition 4 *Let $a, b \in F[x, y_1, \ldots, y_n]$ be polynomials of degree at most d in x, $y_1, \ldots, y_n$. Then there exist polynomials $u, v \in F[x, y_1, \ldots, y_n]$ and $d \in F[y_1, \ldots, y_n]$ respectively, s.t. $au + vb = d$, iff a and b have no common factor of positive degree in x. Furthermore, such a d exists with degree at most $O(d^2)$.*

Discriminant

We are now back to working with polynomials in $F[x]$, for a field F.

Suppose we'd like to know if a polynomial f has a nontrivial divisor g such that g^2 also divides f. We know that over the reals, such g exists if and only if $(f, f') \neq 1$ (here f' is the derivative of f). This is because if $f = g^2h$, then $f' = 2gg'h + g^2h' = g(2g'h + gh')$. Conversely, if $(f, f') = g \neq 1$, then $f = gh$ and $f' = gk$, so $f' = g'h + h'g = gk$, so $g|g'h$, so g and h have a non-trivial common factor t (because $F[x]$ is a unique factorization domain and $\deg(g) > \deg(g')$ and $g' \neq 0$, because $g \neq 1$). Thus, $t|g$ and $t|h$, so $t^2|gh = f$.

But the derivative f' doesn't really make sense over a non-continuous domain such as a finite field or the rationals. However, it can still be syntactically defined: if $f(x) = \sum_{i=0}^{n} f_i x^i$, then $f'(x) = \sum_{i=1}^{n} i f_i x^{i-1}$ (of course, we need to define if_i, because the field F may not contain the integers; we defined it naturally as follows: for $i \in \mathbb{Z}$ and $\alpha \in F$, $i\alpha = \sum_{j=1}^{i} \alpha$). All the expected properties of derivatives can be syntactically verified: $(a+b)' = a' + b'$, $(ab)' = a'b + b'a$ and $(a^2)' = 2aa'$. It would then seem that we could make the same statement about (f, f') over arbitrary fields as we did over the reals and that the same proof as above would go through. Indeed it does, if $f' \neq 0$. The only problem is that, over fields of characteristic $p > 0$, f' can be 0 even if $\deg(f) > 0$ (we used the fact that f' cannot be 0 in the proof above). This will happen if and only if $f = \sum_{i=0}^{n} f_i x^{pi}$. To handle this case we notice that the following conditions hold over a field of size p^k: (1) $a^p + b^p = (a+b)^p$ for every $a, b \in F[x]$, and (2) $a = a^{p^k} = (a^{p^{k-1}})^p$. Thus, we get that $g = \left(\sum_{i=0}^{n} g_i x^i\right)^p = t^p$, where $t = \sum_{i=0}^{n} g_i^{p^{k-1}} x^i$. Thus $t^2 | g$, so $t^2 | f$.

This leads us to the following definition and the following (just proven) theorem.

Definition 5 *The discriminant of a polynomial $f \in F[x]$, denoted $\Delta(f)$ is defined to be* $\mathrm{Res}(f, f')$.

Theorem 6 *Let $f \in F[x]$. The discriminant of f is zero if and only if $g^2 | f$ for some $g \in F[x]$,* $\deg(g) > 0$.

As in the case of the resultant, the discriminant can also be defined over commutative rings. Over unique factorization domains, it retains properties similar to the case of discriminant over fields.

Some Asides

The following statements demonstrate more properties of the resultants.

Proposition 7 *If $a(x), b(x) \in F[x]$ are of the form $a(x) = \prod_i (x - \alpha_i)$ and $b(x) = \prod_j (x - \beta_j)$, then* $\mathrm{Res}(a, b) = \prod_i \prod_j (\alpha_i - \beta_j)$.

To prove the above statement, one expresses the coefficients of $a(x)$ and $b(x)$ explicitly in terms of the α_i's and β_i's. Then by inpecting the matrix $M_{a,b}$ one can see that the determinant consists only of homogenous terms in α_i's and β_j's; and thus the resultant $\mathrm{Res}(a, b)$ is a multivariate polynomial in the α's and β's of total degree mn. Furthermore, $\alpha_i - \beta_j$ must divide this polynomial, since if $\alpha_i = \beta_j$, then the resultant is zero. Thus the resultant is of the form described above.

A similar formula also exists for the discriminant of $f(x)$ if $f(x) = \prod_i (x - \alpha_i)$.

Proposition 8 *If $f(x) \in F[x]$ is given by $f(x) = \prod_{i=1}^{n} (x - \alpha_i)$, then $\Delta(f) = \prod_{i<j} (\alpha_i - \alpha_j)^2$.*

Finally we see a "typical" application of the resultant.

Proposition 9 *Let $a(x, y), b(x, y) \in F[x, y]$ be of degree at most d in x, y. Suppose for at least cd^2 choices of $\alpha \in F$, it is the case that the univariate polynomial $a^x_\alpha(y) \stackrel{\text{def}}{=} a(\alpha, y)$ divides the univariate polynomial $b^x_\alpha \stackrel{\text{def}}{=} b(\alpha, y)$. (Here c is some universal constant.) Suppose further that for at least cd^2 choice of $\beta \in F$ we have that $a^y_\beta(x) \stackrel{\text{def}}{=} a(x, \beta)$ divides $b^y_\beta(x) \stackrel{\text{def}}{=} b(x, \beta)$. Then $a(x, y)$ divides $b(x, y)$.*

To prove the above proposition, assume a does not divide b. By removing common factors, one is left with a and b that are relatively prime. Thus one can consider the resultant of a and b wr.t. y. This must be non-zero and hence it is a low-degree polynomial in x, say $r(x)$. For a specific choice of $x = \alpha$, the resultant of the polynomials a^x_α and b^x_α equals $r(\alpha)$. If α is such that a^x_α divides b^x_α, the α is a root of r. Since r can not have too many roots, the proposition would follow. (But where did we need the conditions on a^y_β dividing b^y_β?)

Factoring Polynomials over Finite Fields

Let's start by looking at a simple problem. How do we factor $x^2 - a$ over an odd prime finite field $GF(p) = \mathbb{Z}_p$, for $a \neq 0$? We know that if it does factor, it will be $(x - \alpha)(x + \alpha)$, for $\alpha^2 = a$.

We also know that, by Fermat's Little Theorem, $\beta^p = \beta$ for all $\beta \in \mathbb{Z}_p$. Therefore, every $\beta \in \mathbb{Z}_p$ is a root of $x^p - x$, therefore $(x - \beta) | (x^p - x)$. Thus, $x^p - x = \prod_{\beta \in \mathbb{Z}_p} (x - \beta)$ (there are no other factors because the right-hand side is degree p). Thus, both $(x - \alpha)$ and $(x + \alpha)$ divide $x^p - x$.

Let $f_1(x) = x^{\frac{p-1}{2}} - 1$ and $f_2(x) = x^{\frac{p-1}{2}} + 1$ Note that $x^p - x = xf_1(x)f_2(x)$. Therefore, $x - \alpha$ divides $f_1(x)$ or $f_2(x)$, and $x + \alpha$ divides $f_1(x)$ or $f_2(x)$. If they divide different ones, then we can use the GCD algorithm on $(x^2 - a)$ and $f_1(x)$ to get $(x - \alpha)$ or $(x + \alpha)$. Note, of course, that the GCD algorithm runs in time polynomial in the degrees of its inputs, so this, if done directly, would take time polynomial in p, which may be to long. Instead, you first compute $g_1(x) = f_1(x) \mod x^2 - a$ by computing $x \mod x^2 - a$, $x^2 \mod x^2 - a$, $x^4 \mod x^2 - a, \ldots$ and then multiplying the necessary ones together to get $x^{\frac{p-1}{2}} \mod x^2 - a$. Then you compute $(g_1(x), x^2 - a) = (f_1(x), x^2 - a)$.

Thus, the only remaining question is how to make it so that $(x - \alpha)|f_1(x)$ and $(x + \alpha)|f_2(x)$ (or vice versa)? Note that $(x - \alpha)|f_1(x) \Leftrightarrow \alpha^{\frac{p-1}{2}} = 1$ and $(x + \alpha)|f_1(x) \Leftrightarrow (-\alpha)^{\frac{p-1}{2}} = -1$. Since for all $\alpha \in \mathbb{Z}_p$, $\alpha^{\frac{p-1}{2}} = 1$ or -1, the condition that interests us is equivalent to $(-1)^{\frac{p-1}{2}} = -1$. This is the case if and only if $\frac{p-1}{2}$ is odd, or $p = 3 \mod 4$.

So, we now have an algorithm for factoring polynomials $x^2 - a$ over finite fields of prime size $p = 3 \mod 4$. (We could have got to this point more quickly by using $\alpha = a^{(p+1)/4}$!) In the next lecture, we will see how to generalize the above algorithm to all fields, and to factor all polynomials!