6.S897 Algebra and Computation March 5, 2012

Lecture 8

Lecturer: Madhu Sudan *Scribe: Eric Miles*

In the last lecture, we saw a method for factoring univariate polynomials over finite fields. Today, we will begin covering the method for factoring bivariate polynomials. Throughout, we will restrict our attention to monic polynomials.

Though the bivariate factoring algorithm applies to polynomials $f \in \mathbb{F}[x, y]$ for a field $\mathbb{F}$, we will view f as an element of $R[x]$ for a commutative ring R. Taking $R = \mathbb{F}[y]$ of course gives $f \in \mathbb{F}[x, y]$, but much of what we will cover is also applicable when $R = \mathbb{Z}$, and thus today's lecture will also be useful when we discuss factoring polynomials over the integers.

One tool used in the analysis of bivariate factoring is the *resultant* $\text{Res}(f, g)$ of ring elements $f, g \in R[x]$. Jumping ahead, this has the property that $\text{Res}(f, g) = 0$ iff f and g share a non-trivial factor. The resultant has many applications beyond factoring polynomials, and in Section 3 we will highlight one of these, namely proving one direction of Bézout's Theorem in the plane.

1 Overview

Let $f \in R[x]$ be a monic polynomial over a commutative ring R; we are mainly interested in the case when $R = \mathbb{F}[y]$ for a finite field $\mathbb{F}$, though we will also mention some aspects of the case $R = \mathbb{Z}$. At a high level, the algorithm for factoring f has four stages.

1. Find an ideal $I \subseteq R$. Recall that an ideal is a subset closed under addition ($\forall a, b \in I : a+b \in I$) and under multiplication with R ($\forall a \in R, b \in I : ab \in I$). If $R = \mathbb{F}[y]$ we can take $I = (y) := \{\alpha \cdot y \mid \alpha \in \mathbb{F}[y]\}$, and if $R = \mathbb{Z}$ we can take $I = (p) := p\mathbb{Z}$ for a prime p.

2. Factor f modulo the ideal I, i.e. write f as $f = f_1 \cdot \ldots \cdot f_k \pmod{I}$ for irreducible f_i. The hope is that this step is "easy", using the results of the previous lectures. Indeed, if $R = \mathbb{F}[y]$ then taking any $f \in R[x]$ modulo the ideal (y) gives a univariate polynomial in x, which we know how to factor. (Factoring f modulo a prime p over $\mathbb{Z}[x]$ is similarly possible.)

3. "Lift" the factors f_i to polynomials $\widetilde{f}_i$ such that $f = \widetilde{f}_1 \cdot \ldots \cdot \widetilde{f}_k \pmod{I^t}$ for a sufficiently large value of t. This uses a technique called Hensel lifting, which will be covered in a future lecture. Note here that I^t is the additive closure of $\{\alpha_1 \cdot \ldots \cdot \alpha_t \cdot \beta \mid \alpha_i \in I, \beta \in R\}$.

4. "Jump" from $\widetilde{f}_1$ to get a polynomial g that divides f in $R[x]$. This step will be the focus of today's lecture, and we will see that it reduces to solving a linear system over $\mathbb{F}$ when $R = \mathbb{F}[y]$. This step is significantly more complicated when $R = \mathbb{Z}$; in this case it reduces to solving a certain lattice problem, and in the next lecture we will see how to solve this problem using the Lenstra-Lenstra-Lovász (LLL) algorithm.

Already, one might be a bit skeptical about this approach. Consider $f = x^p - x + y \in \mathbb{F}_p[x, y]$, which is irreducible over $\mathbb{F}_p$. Taking f modulo the ideal $I = (y)$ gives the polynomial $x^p - x$, and we know from previous lectures that this splits into p distinct linear factors over $\mathbb{F}_p$. However, all factors of f modulo I^t for any $t \geq 2$ must be trivial (i.e. either 1 or f).

More generally, suppose that f has the factorization $g_1 \cdot \ldots \cdot g_\ell$ in $R[x]$. Can f have fewer factors when it is taken modulo I? Can it have more factors? The answer to both questions is, in fact, yes. To see that it can have fewer factors, consider the case when $g_i = \alpha + y \cdot h(x)$ for some $h \in \mathbb{F}[x]$ and $\alpha \in \mathbb{F}$. Then $g_i \pmod{I}$ is a constant, so f will have fewer (non-trivial) factors modulo I.

However this is actually a rare event, and we can circumvent this case by instead using a different ideal $I = (y + \beta)$ for an appropriately chosen $\beta \in \mathbb{F}$.

To see that f can have more factors modulo I, take the irreducible polynomial $f = x^p - x + y$ above which has p factors modulo I. Unlike the case where $f \pmod I$ has fewer factors, we can not simply circumvent this possibility, and there is a potentially one-to-many correspondence between the factors of f and the factors of $f \pmod I$:

$$\begin{aligned} f(x,y) &= \overbrace{g_1(x,y)}\cdot \overbrace{g_2(x,y)}\cdot \ldots \cdot \overbrace{g_\ell(x,y)} \\ f \pmod I &= \overbrace{f_1(x)\cdot \ldots \cdot f_{i_1}(x)}\cdot \overbrace{f_{i_1+1}(x)\cdot \ldots \cdot f_{i_2}(x)}\cdot \ldots \cdot \overbrace{f_{i_{\ell-1}}(x)\cdot \ldots \cdot f_k(x)} \end{aligned}$$

However when we cover Hensel lifting we will see that this state of affairs is acceptable, and that repeatedly lifting f_1 will give an $\widetilde{f_1}$ that has enough information to recover g_1.

2 The "jump" step

We now explain the mechanics behind step 4 in the above algorithm.

Suppose that f splits into factors $g_1 \cdot \ldots \cdot g_\ell$ (unknown to us), and that we have the factorization $f = f_1 \cdot \ldots \cdot f_k \pmod I$. Then this is also a factorization of $\prod_i g_i \pmod I$, and so f_1 is a factor of one of the $g_i \pmod I$; say without loss of generality that it's a factor of g_1. In step 3, we obtain via Hensel lifting the factors $\widetilde{f_1} \cdot \ldots \cdot \widetilde{f_k}$, with the guarantee that $\widetilde{f_1}$ is a factor of $g_1 \pmod{I^t}$ under our assumption that f_1 is a factor of $g_1 \pmod I$. (Note that we have not yet specified an appropriate value of t.) Define $d := \deg_x(\widetilde{f_1})$ to be the x-degree of $\widetilde{f_1}$. Note that $d < \deg_x(f)$ (unless f is irreducible modulo I^t), but $\deg_y(\widetilde{f_1})$ might be very large.

Given this setup, we can now state the Jump Problem.

The Jump Problem. Find two polynomials $g, h \in \mathbb{F}[x,y]$ that satisfy the following conditions.

1. $\deg_x(g) \le d$ and $\deg_y(g) \le d$.
2. $g = \widetilde{f_1} \cdot h \pmod{I^t}$.
3. g has minimal x-degree.

In a moment we'll come to why these polynomials might be useful for us, but first let's focus on solving this problem. One thing to notice is that if (g,h) and (g',h') are two pairs that satisfy condition 2, then their sum $(g+g', h+h')$ also satisfies condition 2. In fact, it turns out that the Jump Problem reduces to simply solving a system of linear equations determined by $\widetilde{f_1}$ and I^t, where the unknowns are the coefficients of g and h. (That this is indeed a linear system relies on the fact that multiplying by $\widetilde{f_1}$ and reducing modulo I^t are both linear operations.) Solving such a system can be done efficiently using basic linear algebra, and we omit the details.

We now explain why a solution to the Jump Problem is useful for us. Recall that we are hoping to find the irreducible polynomial g_1 such that $f = g_1 \cdot h_1$, where here $h_1 := g_2 \cdot \ldots \cdot g_\ell$. The following lemma shows that, given any solution (g,h) to the Jump Problem, we can find a non-trivial factor of f containing g_1 by computing $\gcd(f,g)$. (If f is irreducible, this gcd will give f.)

Lemma 1 *If (g_1, h_1) is a solution to the Jump Problem with g_1 irreducible, and (g_2, h_2) is any other solution, and if $t > d^2$, then $g_1 | g_2$.*

(Note that this also specifies the value of t we need to choose in step 3.) We will not prove this lemma today. Instead we will introduce the *resultant*, which is a generally useful tool that in particular will help us prove this lemma.

Before doing so, we briefly remark on how the above discussion differs in the case when $R = \mathbb{Z}$. Recall that when $R = \mathbb{Z}$, we take the ideal $I = p\mathbb{Z}$ for a prime p. So, step 2 factors $f \in \mathbb{Z}[x]$ as $f = f_1 \cdot \ldots \cdot f_k \pmod{p}$, and then step 3 lifts these to $f = \widetilde{f_1} \cdot \ldots \cdot \widetilde{f_k} \pmod{p^t}$. As before, we know that $\widetilde{f_1}$ has small x-degree $d < \deg_x(f)$, and we again try to find a pair (g, h) such that $g = \widetilde{f_1} \cdot h \pmod{p^t}$. The only difference now is that, instead of requiring g to have minimal x-degree, we require it to have coefficients with absolute value bounded by 2^{b^2}, where 2^b is a bound on the magnitude of f's coefficients. This change to the Jump Problem means that it no longer reduces to solving a linear system, but we will see in the next lecture that it reduces to finding a short basis in a certain lattice, and that the LLL algorithm solves this problem efficiently.

3 The resultant

In this section we introduce the resultant, an algebraic tool that will aid in the proof of Lemma 1. To start, consider the following problem:

> Given two polynomials $A = \sum_{i=0}^{k} a_i x^i$ and $B = \sum_{i=0}^{\ell} b_i x^i$ in $R[x]$,
> decide if A and B have a common non-constant factor.

The resultant $\mathrm{Res}_x(A, B)$ solves this problem. We will prove that it has the following properties.

1. $\mathrm{Res}_x(A, B) \in R$.
2. $\mathrm{Res}_x(A, B)$ is a polynomial in the coefficients $\{a_i\}_{i \le k}$, $\{b_i\}_{i \le \ell}$.
3. $\mathrm{Res}_x(A, B)$ is contained in the ideal generated by (A, B).
4. $\mathrm{Res}_x(A, B) = 0$ if and only if A and B have a common non-constant factor.

Note that we use the subscript x to indicate the variable under consideration. If we are working in the ring $R = \mathbb{F}[y]$, as we will below, then the a_i and b_i coefficients are actually polynomials in y.

So, how is the resultant defined? $\mathrm{Res}_x(A, B)$ is the determinant of the following $(k + \ell) \times (k + \ell)$ matrix, known as the *Sylvester matrix* associated with A and B.

$$M(A,B) = \begin{bmatrix} a_0 & 0 & \cdots & 0 & b_0 & 0 & \ldots & 0 \\ a_1 & a_0 & & 0 & b_1 & b_0 & & 0 \\ a_2 & a_1 & & \vdots & \vdots & b_1 & & \vdots \\ \vdots & \vdots & \ddots & a_0 & b_\ell & \vdots & \ddots & 0 \\ a_k & a_{k-1} & & a_1 & 0 & b_\ell & & b_0 \\ 0 & a_k & & a_2 & 0 & 0 & & b_1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_k & 0 & 0 & \cdots & b_\ell \end{bmatrix} \qquad \mathrm{Res}_x(A,B) := \det(M(A,B))$$

This already establishes properties 1 and 2 above, though we will look more closely at the second in a moment. But first, a natural question: where does $M(A, B)$ come from? Its motivation can be found in the proof of the following lemma, which establishes property 4.

Lemma 2 *Let $G := \gcd(A, B)$. Then, G is non-constant if and only if $\det(M(A, B)) = 0$.*

Proof Assume that G is non-constant. Then, $A \cdot (B/G) + B \cdot (-A/G) = 0$, and thus there exist two non-zero polynomials $C := \sum_i c_i x^i$ and $D := \sum_i d_i x^i$ such that $AC + BD = 0$, $\deg(C) < \deg(B)$, and $\deg(D) < \deg(A)$. Then, defining the (column) vector $v = (c_0, \ldots, c_{\ell-1}, d_0, \ldots, d_{k-1}) \neq 0$, we have $M(A, B) \cdot v = AC + BD = 0$ and thus $\det(M(A, B)) = 0$. This argument also holds in the other direction, i.e. if $\det(M(A, B)) \neq 0$ then there is no such $v \neq 0$ and so G must be constant. ■

We now note a few other facts about the resultant. Applying the following general lemma to our matrix shows that the vector $(\mathrm{Res}_x(A, B), 0, \ldots, 0)$ is in the column span of $M(A, B)$, which establishes property 3.

Lemma 3 *For all $M \in R^{n \times n}$, the vector $(\det(M), 0, \ldots, 0)$ is in the column span of M.*

Proof M can be put in lower-triangular form by performing only column operations. Letting $\overline{M}$ denote the triangularized matrix, we have $\det(M) = \prod_{i \leq n} \overline{M}_{ii}$. Finally, observe that the vector $\left(\prod_{i \leq n} \overline{M}_{ii}, 0, \ldots, 0\right)$ is in the column span of any triangular matrix $\overline{M}$. ■

In the case when $R = \mathbb{F}[y]$, the following lemma bounds the y-degree of $\mathrm{Res}_x(A, B)$.

Lemma 4 *If $A, B \in \mathbb{F}[x, y]$ have total degree k and ℓ respectively, then $\mathrm{Res}_x(A, B) \in \mathbb{F}[y]$ has degree at most $k\ell$.*

Proof This is essentially a counting argument. Consider the degree of the (i, j)th element of $M(A, B)$:

$$\deg(M(A, B)_{ij}) \leq \begin{cases} k - i + j, & \text{if } j \leq \ell \\ j - i, & \text{if } j > \ell. \end{cases}$$

Therefore for every permutation $\sigma : [k + \ell] \to [k + \ell]$, $\deg\left(\prod_j M(A, B)_{\sigma(j),j}\right) \leq k\ell$, and so $\mathrm{Res}_x(A, B) = \det(M(A, B))$ is a sum of degree $\leq k\ell$ polynomials. ■

We conclude by showing how the resultant can be used to prove one direction of Bézout's Theorem in the plane.

Theorem 5 *If $A, B \in \mathbb{F}[x, y]$ have total degree at most k and ℓ respectively, and they share more than $k\ell$ common zeros, then they have a common non-constant factor.*

Proof We will show that if A and B have $> k\ell$ common zeros, then $\mathrm{Res}_x(A, B) = 0$. Suppose that $(\alpha_1, \beta_1), \ldots, (\alpha_{k\ell+1}, \beta_{k\ell+1})$ are the common zeros. We know that $\mathrm{Res}_x(A, B)$ is in the ideal generated by A and B, so it must vanish on each of the β_i. Because $\mathrm{Res}_x(A, B)$ has y-degree $\leq k\ell$ by Lemma 4, if each of the β_i are distinct then $\mathrm{Res}_x(A, B)$ must be identically zero. Of course, the assumption that the β_i are distinct is not justified. However, if we work over a large enough extension field $K \supseteq \mathbb{F}$, and perform the following linear transformation for a random $\theta \in K$

$$(\alpha_i, \beta_i) \longmapsto (\alpha_i,\ \beta_i + \theta \cdot \alpha_i)$$

then with non-zero probability the new β_i will all be distinct, which again gives $k\ell + 1$ distinct points on which $\mathrm{Res}_x(A, B)$ vanishes. ■