Factoring Rational Polynomials over the Complex Numbers

Chanderjit Bajaj [*] John Canny [†] Thomas Garrity [‡]
Joe Warren [§]

February 1989

Abstract

NC algorithms are given for determining the number and degrees of the factors, irreducible over the complex numbers $\mathbf{C}$, of a multivariate polynomial with rational coefficients and for approximating each irreducible factor. NC is the class of functions computable by logspace-uniform boolean circuits of polynomial size and polylogarithmic depth. The measures of size of the input polynomial are its degree d, coefficient length c, number of variables n. If n is fixed, we give a deterministic NC algorithm. If the number of variables is not fixed, we give a random (Monte-Carlo) NC algorithm in these input measures to find the number and degree of each irreducible factor.

After reducing to the two variable, square-free case, we apply the classical algebraic geometry fact that the absolute irreducible factors of $(P(z_1, z_2) = 0)$ correspond to the connected components of the real surface (or *complex* curve) $P(z_1, z_2) = 0$ minus its singular points. In finding the number of connected components of the surface $P = 0$,

[*]Department of Computer Science, Purdue University. Supported in part by ARO Contract DAAG29-85-C0018 and ONR contract N00014-88-K-0402

[†]Computer Science Division, Berkeley. Supported in part by a David and Lucille Packard Fellowship

[‡]Department of Mathematics, Rice University

[§]Department of Computer Science, Rice University. Supported in part by NSF grant IRI 88-10747

the surface is projected to the the z_2- plane. The singular points of $P(z_1, z_2)$ lie over the projection's critical values. The inverse image of a grid isolating the critical values in the z_2- plane lifts to a one dimensional real curve skeleton on the surface $(P = 0)$ whose number of connected components is precisely the number of connected components of $P = 0$ minus its singular points. The connectivity of this curve skeleton is constructed symbolically using Sturm sequences associated with the various polynomials defining these maps. Given the number of irreducible factors and their degree, the actual factors can be reconstructed using the recent result of Neff [22] on finding zeroes of one variable polynomials in NC.

1 Introduction

Factoring polynomials is a basic problem in symbolic computation with applications as diverse as theorem proving and computer-aided design. Our goal is to approximate the factors, irreducible over the complex numbers, of a multivariable polynomial with rational coefficients in deterministic NC with respect to the polynomial's degree and coefficient size, assuming that the number of variables is fixed. Further if the number of variables is not fixed, we will find the number of irreducible factors and each of their degrees in random NC with respect to the polynomial's degree, coefficient size and number of variables. The key is that this is a topological problem, not an algebraic one. We will show that the number of irreducible factors is exactly equal to the number of connected components of a certain semi-algebraic set.

Since factoring is held as a touchstone problem in computer algebra, there have been a number of recent successes in factoring various types of polynomials over different fields. Methods for factoring polynomials with rational coefficients over the rational numbers are well-known. Kaltofen [16] and Lenstra, Lenstra and Lovasz [20] establish that factoring polynomials in a fixed number of indeterminates over the field of rational numbers $\mathbf{Q}$ is in polynomial time.

For factoring rational polynomials over $\mathbf{C}$, Noether [23], Davenport and Trager [6] and Heints and Sieveking [13] each give methods that require time exponential in the degree of the input polynomial. DiCrescenzo and Duval [7] and Duval [8] give geometric methods of factorization based on algebraic geometry. Kaltofen [15] describes an NC method for testing whether

a rational polynomial is irreducible over $\mathbf{C}$. The method involves computing approximate roots and their corresponding minimum polynomials. The first polynomial time algorithm for factoring over $\mathbf{C}$ seems to have been given by Chistov and Grigoryev [4]. Until recently, it has been an open problem whether approximating the factors, irreducible over $\mathbf{C}$, of a rational polynomial is in NC. Kaltofen [18], using techniques quite different from ours, has also derived an algorithm for approximating these factors in NC.

Recall that NC is the class of functions computable by log-space uniform Boolean circuits of polynomial size and polylogarithmic depth. Thus the running time of an NC algorithm will be polylogarithmic, allowing a polynomial number of processors that work in parallel. Given a polynomial P with rational coefficients, the input size is measured by number of variables n, degree d, coefficient size c, and number of non-zero coefficients s. We show that the general problem of computing number and degrees of the factors is in random NC in these measures, in the Monte-Carlo sense (definitely fast, probably correct). If the number of variables is fixed, or if the polynomial P is dense, we give a deterministic NC solution for also approximating the irreducible factors. Finally, if the polynomial is represented as a straight-line program of length p our algorithm runs in random NC plus the time to evaluate the polynomial at an integer point. By the parallelization result of Valiant et al. [27], any straight-line program of size p and degree d can be converted into an equivalent program of polynomial size, and polylogarithmic depth in d and p, which can be evaluated in NC. Although this result applies to the real number model, it extends easily to bit complexity since every node of the SLP represents a polynomial in the input variables and constants. If we have a bound on the degree of these polynomials, this bounds also the size of any intermediate coefficient when we evaluate the program numerically.

However, the conversion to a low-depth straight-line program is itself not in NC, and seems intrinsically sequential because of constant evaluation which is P-complete. So we cannot run our algorithm in random NC for straight-line program polynomials unless we are given a program of low depth.

The paper can be divided into two main parts. In the first part, in sections two and three, we show how to find the number and the degree of the irreducible factors of a two variable, square-free polynomial with rational coefficients. In the second main part, in sections four and five, we show how to reduce the general case of a multi-variable polynomial to a two variable, square-free polynomial, and then how to approximate the actual irreducible

factors.

In section two, we recall the classical algebraic geometry fact that the number of factors of a two variable, square-free polynomial $P(z_1, z_2)$ is precisely the number of connected components of the real two dimensional surface $P = 0$ minus the singular points of P. This is the key, allowing us to translate the original algebraic problem of factoring to the topological one of finding connected components of a special semi-algebraic set. Section three, the technical heart of the paper, finds the number of these connected components. The two dimensional surface $P(z_1, z_2) = 0$, lying in $\mathbf{C^2}$ or equivalently $\mathbf{R^4}$, can be projected to the z_2 plane, with a finite number of critical values. The singular points of the surface lie over critical points of the projection. In section 3.1, the projection is described. In section 3.2, we create a grid of horizontal and vertical lines into the z_2- plane with the property that inside each box of the grid there is at most one critical point of the projection. The inverse image of this grid on the surface $P(z_1, z_2) = 0$ forms a curve skeleton. By theorems three and four, the number of connected components of this curve skeleton will be precisely the number of connected components of $P(z_1, z_2) = 0$ minus its singular points. Since this curve skeleton is a graph, if we can compute its adjacency matrix in NC, we can find the number of connected components in NC. Section 3.3 shows how to find this matrix by using sign sequences of various Sturm sequences, all of which, by theorem six, can be computed in NC. This will give the number of irreducible factors of a two variable square free polynomial with rational coefficients. We will see that the degree of each factor has already been determined, by applying Bezout's theorem.

In section four, we show how to the reduce the problem of factoring a multi-variable polynomial with rational coefficients to that of factoring a square-free, two variable polynomial. All we do is note that the proof of Bertini's theorem in ([21]) actually describes an algorithm that runs in NC. Finally, in section five, using the recent wonderful result of Neff [22] that the roots of a one variable rational polynomial can be approximated in NC, we show how to actually approximate the irreducible factors.

2 Connectivity and Factorization

2.1 Preliminaries

Let $P_i(z_1, \ldots, z_n) \in \mathbf{C}[z_1, \ldots, z_n]$ for $i = 1, \ldots, k$ be polynomials with complex coefficients in n variables. Let $V(P_1, \ldots, P_k)$ denote the set of common zeros of these polynomials in $\mathbf{C}^n$

$$V(P_1, \ldots, P_k) = \{z \in \mathbf{C}^n | P_i(z) = 0, \quad i = 1, \ldots, k\}$$

This is an example of an *algebraic* set. For a single polynomial P, the set $S = V(P)$ is called a *hypersurface*. A hypersurface S is said to be *irreducible* if it is the zero set of a polynomial $P(z_1, \ldots, z_n)$ irreducible over $\mathbf{C}$. More generally, an algebraic set is irreducible if it cannot be expressed as a finite union of proper algebraic subsets. An irreducible algebraic set is called a *variety*.

For the rest of the paper, we assume that P is square-free (irreducible factors have multiplicity one). Note that if the original P is not square-free, we may compute the square-free part of P by computing

$$P/GCD(P, \frac{\partial P}{\partial z_1})$$

where P is monic in z_1. This computation may be performed in NC using greatest common divisor algorithm of [1].

The key observation in section 2.2 will be that there is a fundamental relationship between the *singular* points of an algebraic set and its irreducible components.

Definition Let $S = V(P)$ be a hypersurface with P a square-free polynomial . The set of *singular points* of S, denoted $Sing(S)$, is defined by

$$Sing(S) = S \cap V(\frac{\partial P}{\partial z_1}, \ldots, \frac{\partial P}{\partial z_n}). \tag{1}$$

For example, an irreducible algebraic plane curve has at most a finite number of singular points. More generally, the singular set can be defined for any algebraic set, but we will not give a definition here. Intuitively, the singular points of an algebraic set are the points where the set is not smooth (smooth points have neighborhoods diffeomorphic to some $\mathbf{C}^k$) or where tangent lines are ill-defined.

2.2 Topology of Zero Sets of Reducible Polynomials

This section is the key to translating the algebraic problem of factoring to the topological problem of counting connected components.

Removing the singular set from an algebraic set may split it into several connected components. Here connectivity is equivalent to pathwise connected in the usual (metric) topology. As the following theorems show, these components correspond exactly to the irreducible components of the curve.

Theorem 1 *The set S is irreducible if and only if $S - Sing(S)$ is connected.*

Theorem 2 *The irreducible components of set S are exactly the closures of the connected components of $S - Sing(S)$.*

Both theorems are extremely classical and are consequences of the next two lemmas.

Lemma 1 *Let the set S have distinct irreducible components $S_1, S_2, \ldots, S_k$. Then for any i and j, $S_i \cap S_j \subseteq Sing(S)$.*

For hypersurfaces, this is a straightforward calculation. For the general case, this is theorem 6 in chapter two, section 2 of [26].

Lemma 2 *If S is irreducible, and Y is any proper algebraic subset of S, then $S - Y$ is connected.*

This is Corollary (4.16) of [21].

3 Computing Connected Components

We have reduced the problem of finding the number of irreducible factors of $P(z_1, z_2)$ to the problem of counting the number of connected components of the surface $V(P) - Sing(V(P))$. In the first part of this section, we discuss an algorithm for counting the number of connected components. We then discuss how the algorithm can be executed in parallel.

3.1 Projecting to the Plane

In this subsection, we treat the algebraic set $V(P(z_1, z_2))$ as a branched cover of the z_2 plane, showing that there will only be a finite number of critical values (which will be defined in a minute) and, more importantly, that the singular points of P must lie over critical values. In sections 3.2 and 3.3, we will construct a grid, isolating the critical values, in the z_2 plane whose inverse image on $V(P)$ will be a graph with the same number of connected components as $V(P) - Sing(V(P))$. This reduces the problem to constructing the adjacency matrix of this graph.

We express the complex coordinates z_1 and z_2 in terms of their real and imaginary parts: $z_1 = x_1 + y_1\mathbf{i}$ and $z_2 = x_2 + y_2\mathbf{i}$. The polynomial $P(z_1, z_2) = P(x_1, y_1, x_2, y_2)$ can also be expressed in terms of its real and imaginary parts: $P(x_1, y_1, x_2, y_2) = P_1(x_1, y_1, x_2, y_2) + P_2(x_1, y_1, x_2, y_2)\mathbf{i}$. Thus, $V(P)$ can be expressed as the two dimensional real surface $V(P_1, P_2)$ in $\mathbf{R^4}$. Let $\pi : \mathbf{R}^4 \to \mathbf{R}^2$ be the projection map that takes those points (x_1, y_1, x_2, y_2) lying on $V(P_1, P_2)$ and maps them to (x_2, y_2). By a change of variables, we can assume that $P(z_1, z_2)$ has a z_1^d term, where d is the degree of P, implying that P does not have any factors univariate in z_2. Thus, π must be finite-to-one everywhere.

Definition If $F : \mathbf{R}^n \to \mathbf{R}^m$ is a differentiable map, $p \in \mathbf{R}^n$ is a *critical point* of F if the Jacobian dF of F is not surjective at p. The image of a critical point is a *critical value.*

In complex algebraic geometry, the critical points are called *ramification points* and the critical values *branch points.* A point which is not a critical point is called a *regular point*, and the preimage of a *regular value* consists of regular points only.

This projection map has only a finite number of critical points.

Lemma 3 *The projection map π from the surface $V(P_1(x_1, y_1, x_2, y_2), P_2(x_1, y_1, x_2, y_2))$ to the (x_2, y_2) plane has only a finite number of critical points.*

This lemma can be stated more precisely as follows:

Lemma 4 *The critical points of the projection map π are the points in the intersection of $V(P)$ and the surface $\frac{\partial P_1}{\partial z_1}$.*

The proofs of both lemmas are contained in sections 4 and 5 in chapter two of [19]. The critical values of the projection map are the zeros of the one variable polynomial $R(z_2) = Res_{z_1}(P, \frac{\partial P}{\partial z_1})$. This expression is frequently referred to as the *discriminant.*

Note that the singular points of the polynomial P, the points where both partial derivatives vanish, must be contained in the critical points.

3.2 Reduction to Curve Skeleton

In this subsection we reduce the problem of finding the number of connected components of $V(P(z_1, z_2)) - Sing(V(P))$, which is the number of irreducible factors of P, to finding the number of connected components of a real one dimensional curve skeleton (or graph) on the surface $V(P) - Sing(V(P))$.

In section 3.1 we showed that the projection map π from $V(P)$ to the z_2 or (x_2, y_2) plane has only a finite number of critical values. Let G be a grid of a finite number of horizontal and vertical lines in the (x_2, y_2) plane, so that each critical value is in at most one cell of the grid. For now, we assume such a grid exists. We will give more details about how to construct such a grid later in this section. For now, the key point is that G isolates the critical values of the projection map.

Let K be the inverse image of the grid G on the surface $V(P)$. Note that K is one dimensional and, since no critical values lie on G, lies on $V(P) - Sing(V(P))$. K can be interpreted as defining a graph whose vertices are those points on K lying over the vertices of the grid G. Two vertices in this graph are adjacent if their corresponding points on K are connected.

The following two theorems will show that the number of connected components of the curve skeleton K is the same as the number of connected components of $V(P) - Sing(V(P))$. Theorem 3 will also yield the degree of of each irreducible factor of $P(z_1, z_2)$

Theorem 3 *The number of vertices of the curve skeleton K, over any vertex of the grid G, in each connected component of $V(P) - Sing(V(P))$ is exactly the degree of that component.*

Proof Let $P(z_1, z_2) = \Pi Q_i(z_1, z_2)$, where each Q_i is an irreducible factor of P. Each connected component of $V(P) - Sing(V(P))$ is of course $V(Q_i) - Sing(V(P))$, for some i.

A vertex of the grid G is a point a in the z_2 plane. The inverse image of this vertex is given by the zeroes of the one variable polynomial $P(z_1, a)$. The inverse image in each component are the zeroes of $Q_i(z_1, a)$. By the Fundamental Theorem of Algebra, each component will have exactly degree of Q_i points, which is also the degree of the component. We do not have to worry about multiple roots of these polynomials since the the grid G does not pass through any critical values. **QED**

Theorem 4 *Any path in $V(P) - Sing(V(P))$ connecting two points in K is homotopic (i.e. can be continuously deformed) to a path in K.*

Proof Let σ be a path in $V(P) - Sing(V(P))$ connecting two points of K. By slightly deforming σ, we can assume that the projection of σ by π misses any critical values in the z_2 plane. In each cell of the grid G, deform the path $\pi(\sigma)$ to the actual grid G, so that the area swept out by the deformation does not contain any critical values. This is possible since each cell will contain at most one critical value. Since the area swept out by the deformation does not contain a critical value, the deformation can be lifted to $V(P) - Sing(V(P))$. Thus the path σ can be continuously deformed to a path on the skeleton K. **QED**

We next describe the actual construction of the grid G. In the last section we showed that the critical points of the projection map π are the points in $V(P, \frac{\partial P}{\partial z_1})$. Then the critical values are the zeroes of the one variable polynomial,

$$R(z_2) = Res_{z_1}(P, \frac{\partial P}{\partial z_1}), \tag{2}$$

where $Res_{z_1}(P, Q)$ is the resultant of the polynomials of P and Q treated as one variable polynomials in z_1. The edges of G are chosen to be parallel to the x_2 and y_2 axes. The vertical edges are the lines $x_2 = v_i$ with the $(v_0 < v_1 < \cdots)$ real constants chosen so that the open interval (v_i, v_{i+1}) contains at most one of the distinct real components of the complex zeroes of $R(z_2)$. The horizontal edges are the lines $y_2 = h_i$ with the $(h_0 < h_1 < \cdots)$ real constants chosen so that the open interval (h_i, h_{i+1}) contains at most one of the distinct imaginary components of the complex zeroes of $R(z_2)$. Figure 1 illustrates this situation where $R(z_2) = (z_2 - 1)(z_2 + \mathbf{i})(z_2 - \mathbf{i})$.

The lines of G form rectangular cells in the z_2 plane, intersecting in vertices. Note that each cell in the grid contains at most one critical value.

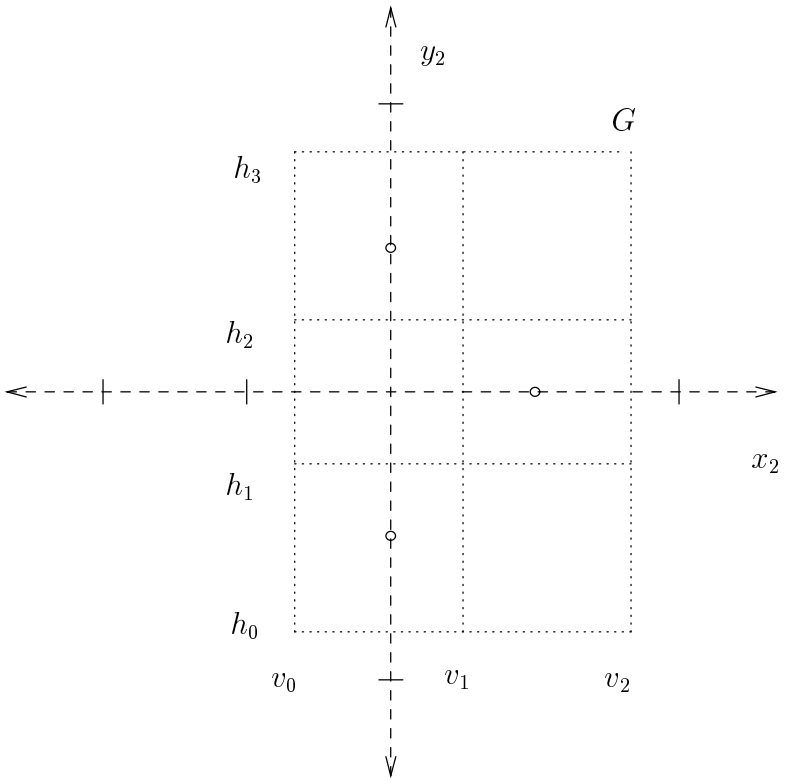

Figure 1: A grid plane whose cells contain at most one critical point.

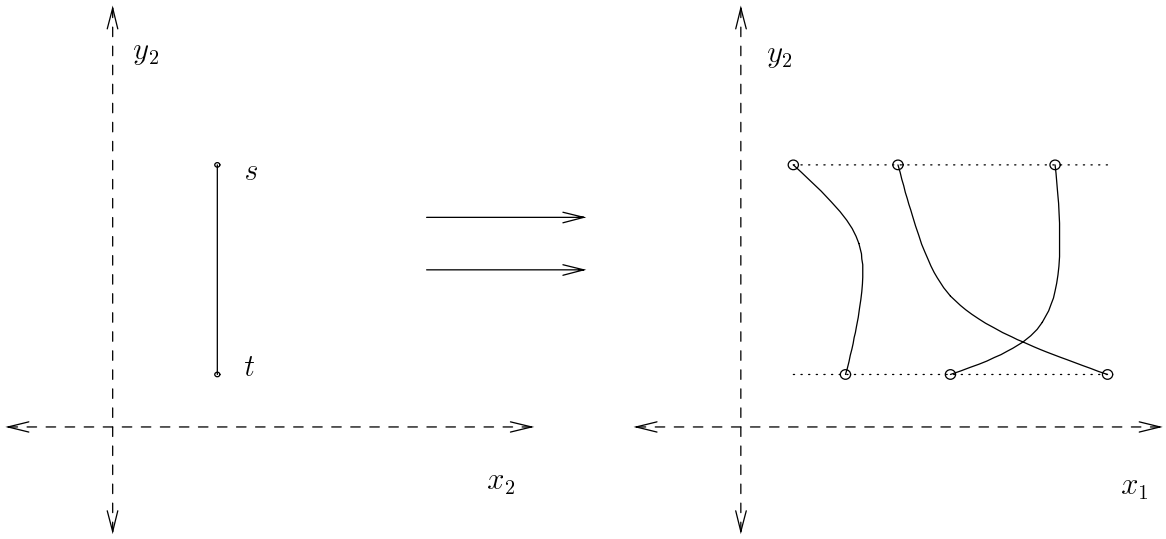

Figure 2: Curve segments on S joining vertices of K

The grid G may now be used to construct the curve skeleton K directly on $V(P)$. In fact K, the inverse image of G under the projection π onto the z_2 plane, is the one dimensional curve $V(P) \cap \pi^{-1}(G)$. As described earlier, the vertices of this graph are the points on $V(P)$ lying over each vertex (v_k, h_l) in G. These points are the complex roots of the univariate polynomial $P(z_1, v_k + \mathbf{i}h_l)$. The edges of the graph correspond to algebraic curve segments of K. Figure 2 illustrates three curves segments over two vertices s and t of the grid G, adjacent on a vertical grid line. The curve segments have been projected onto the x_1y_2 plane.

Thus to find the number of connected components of $V(P) - Sing(V(P))$, we must find the number of connected components of the graph K. To determine the connectivity of K, we need only the adjacency information between points of K, not the actual curve segments. We will next describe a fast parallel method for computing this adjacency information.

3.3 Construction of Curve Skeleton

In this section we show how to construct the adjacency matrix for the curve skeleton K in NC with respect to the degree of the polynomial and the coefficient size. The key will be the symbolic representation by sign sequences

of roots of various polynomials, allowing us to keep track of the order of the roots, which will be needed for the matrix.

In section 3.3.1, we quickly review Sturm sequences for one variable polynomials and then give a generalization for multi-variable polynomials. In section 3.3.2, we show how to explicitly construct the adjacency matrix of the curve skeleton.

3.3.1 Sturm Sequences

Sturm sequences are classical. Let $p(x)$ be a one variable polynomial. Consider the following sequence $p_0(x), ..., p_n(x)$ of polynomials:

$$\begin{array}{c} p_0 = p \\ p_1 = dp(x)/dx \\ \vdots \quad \vdots \\ p_k = q_{k-1}p_{k-1} - p_{k-2} \\ \vdots \\ p_n \end{array} \tag{3}$$

where p_k is simply the negative of the remainder obtained by dividing p_{k-2} by p_{k-1}. Since $p(x)$ is a polynomial, the last term p_n must be a constant. If $p(x)$ is square-free, p_n must be nonzero. Sturm sequences can be computed in NC [1].

The importance of Sturm sequences lies in that they provide an easy way of determining how many real roots a polynomial has between two points.

Theorem 5 *Let $p(x)$ be a univariate real polynomial with Sturm sequence $(p_0(x), \ldots, p_n(x))$. Let a and b be real numbers that are not roots of $p(x)$. Then the number of real roots of $p(x)$ between a and b is equal to the number of sign changes in the sequence $(p_0(a), \ldots, p_n(a))$ minus the number of sign changes in the sequence $(p_0(b), \ldots, p_n(b))$.*

The proof can be found in many places, such as [14, Chapter 6].

We will also need the following multivariable version of Sturm sequences. Let Σ be a collection of rational polynomials $(p_1, ..., p_m)$ in n variables.

Theorem 6 *Let $p_1(x_1, \ldots, x_n) = 0$, ..., $p_m(x_1, \ldots, x_n) = 0$ be a system of rational coefficient polynomial equations having a finite number of solution*

points. Denote the l real solution points not at infinity as $\alpha_j \in \mathbf{R}^n, j = 1, \ldots, l$. Let $q_1(x_1, \ldots, x_n), \ldots, q_k(x_1, \ldots, x_n)$ be a set of polynomials. Then the set of sign sequences of $q_1(\alpha_j), \ldots, q_k(\alpha_j), j = 1, \ldots, l$ can be computed in NC if m is fixed.

This theorem is a corollary of Lemma 2.4 in [3].

3.3.2 Parallel Adjacency Calculation

We now discuss how to compute the grid, G, in the z_2 plane and the adjacency information for K in NC with respect to the degree of the original polynomial and the coefficient size.

Let $R(z_2)$ be the polynomial whose zeroes are the critical values of the projection map from $V(P)$ to the z_2 plane, defined by equation (2). Without loss of generality, we assume that $R(z_2)$ is squarefree (if not, make it so). We write $R(z_2)$ in terms of its real and imaginary parts:

$$R(x_2, y_2) = R_1(x_2, y_2) + \mathbf{i}R_2(x_2, y_2).$$

The complex zeroes of R are at the simultaneous real zeroes of R_1 and R_2. Let

$$\begin{aligned} U(x_2) &= Res_{y_2}(R_1, R_2), \\ H(y_2) &= Res_{x_2}(R_1, R_2). \end{aligned} \tag{4}$$

The real zeroes of U contain the x_2-coordinates of the critical values and the zeroes of H contain the y_2-coordinates. Again, we ensure that U and H are squarefree.

Since U is squarefree, the solutions v_i to the equation

$$\frac{dU(x_2)}{dx_2} = 0 \tag{5}$$

generate vertical lines that separate the critical points. Likewise the solutions h_i to the equation

$$\frac{dH(y_2)}{dy_2} = 0 \tag{6}$$

generate horizontal lines that separate the critical points. Finally, if A is a constant so that all roots of both U and H are greater than $-A$ and less

than A, then the grid G will consist of the lines from equations (5) and (6) and

$$x_2 = \pm A$$
$$y_2 = \pm A.$$

This gives a symbolic description of G. We next use this description of the grid with Sturm sequences to compute the adjacency information for graph of the curve skeleton K.

We describe a method for computing this adjacency information in the x_2 direction in G. The y_2 direction is similar. Let (v_i, h_j) and (v_i, h_{j+1}) be two adjacent vertices in G. These vertices lie on the grid line $x_2 = v_i$. Over each of these vertices lies d points in $V(P)$. These points form the vertices of K. In $\mathbf{R^4}$, the intersection of $x_2 = v_i$ and $V(P)$ define d algebraic curve segments in K. These curve segments form the edges in K, joining pairs of vertices in K, each lying over a distinct grid vertex.

We do not attempt to explicitly construct and follow the curve segments. Instead, we symbolically compute the adjacency information. Project $(V(x_2 - v_i) \cap V(P))$ onto the x_1y_2-plane via resultants. As shown in the next section, this projection can be chosen so that only nodal singularities are introduced into the curve. To determine adjacency information, we need only locate and detect the relative position of these nodes with respect to the vertices of K, since at each node the order of vertices changes. Since we are only interested in relative order of these points, these calculations can be done via Sturm sequences. For example, in figure 3, denote the endpoints of the segment (v_i, h_j) and (v_i, h_{j+1}) by s and t. Over s, assume that there are four vertices s_1, s_2, s_3, and s_4 in K . Likewise over t there are four vertices t_1, t_2, t_3 and t_4 . The projected curve segments link the s_i to the t_j and the position of the nodes determines which s vertices are connected to which t vertices.

Specifically, consider the three polynomials.

$$\begin{aligned} T(x_1, x_2, y_2) &= Res_{y_1}(P_1, P_2). \\ N(x_2, y_2) &= Res_{x_1}(T, \tfrac{\frac{dU(x_2)}{dx_2}}{\partial x_1}). \end{aligned} \tag{7}$$

$V(T)$ is the projection of $V(P)$ to x_1, x_2, y_2 space. Allowing x_1 and y_2 to be free variables for $\frac{dU(x_2)}{dx_2}$, the intersection of $V(\frac{dU(x_2)}{dx_2})$ with $V(P)$ yields curves

lying in planes parallel to the x_1y_2 plane and through the vertical grid lines.

Now allow y_2 to be a free variable for $\frac{dU(x_2)}{dx_2}$, the points on $V(N, dU/dx_2)$, restricted to the line $x_2 = v_i$, are the images of the nodes of $V(T)$ projected to the y_2 axis. Thus, allowing x_1 to be a free variable, $V(N, dU/dx_2)$ consists of lines in the x_1y_2 plane, parallel to the x_1 axis, containing nodes of the projected plane curve (the dotted horizontal lines in figure 3).

Compute the sign sequences of the following polynomials:

- The Sturm sequence of dU/dx_2.
- The Sturm sequence of dH/dy_2.
- The Sturm sequence with respect to y_2 of $N(x_2, y_2)$.
- The Sturm sequence with respect to x_1 of $\frac{\partial T}{\partial x_1}$.

at the common zeros of the system (7). By theorem 6, these sign assignments can be computed in NC with respect to the size of the input polynomials .

To compute adjacencies for K we proceed as follows: As y_2 increases, the number of sign alternations of the Sturm sequence for dH/dy_2 increases monotonically. We first sort all the sign assignments according to the number of sign alternations in this Sturm sequence within each sign assignment. This partitions all the zeros of (7) into classes according to y_2 coordinate. Each of these classes provides adjacency information for a particular slice $y_2 = h_i$.

Next we sort within each class according to number of sign alternations of the Sturm sequence of dU/dx_2. This gives us a collection of classes which lie on the same horizontal grid segment between two adjacent vertical grid lines.

Then sort within classes according to number of sign alternations of the Sturm sequence of $N(x_2, y_2)$. The sign assignments with a zero correspond to the image in the (x_2, y_2) plane of the nodes of the curve seqments.

Finally, we sort the sign assignments according to number of alternations of the Sturm sequence of $\partial T/\partial x_1$. This orders the points with the same x_2, y_2-coordinates along lines parallel to the x_1 axis . One of these sign assignments will have a zero assignment to the polynomial $\partial T/\partial x_1$, and this is the sign assignment of the node point itself. From the position of this sign assignment in the ordering, we infer the relative position of the node point along the dotted line and therefore among the branches of the curve in K. In figure 3, we are ordering the points along the dotted lines.

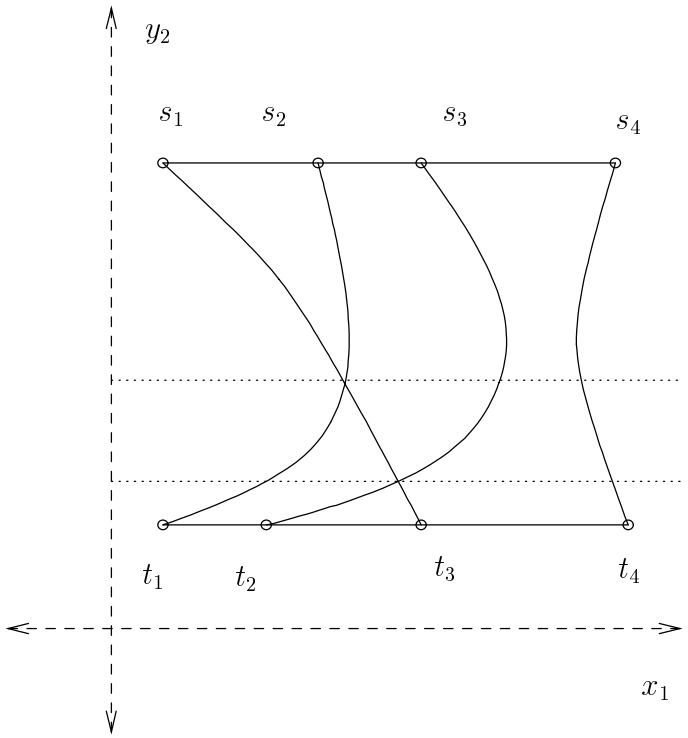

Figure 3: Effect of nodes on adjacency calculations

To generate the graph K, we label the d vertices of K over a given grid point with $1, \ldots, d$. These labels come from the x_1 ordering of the corresponding points in K. Each node can be represented as a permutation (an exchange of two adjacent elements) of the indices of the curve branches that cross at the node. To determine the permutation as we move in y_2 past k nodes, we compose the permutations of the nodes. The composition can be done in NC by composing adjacent (in y_2 ordering) permutations, then composing adjacent pairs of these, etc. The final permutation gives the change in ordering from one grid point to the next, and provides the d edges joining corresponding vertices of K.

One performs similar calculations to compute adjacency information in the horizontal direction.

3.3.3 Projections Introducing only Nodal Singularities

We need to prove the assumption used in the section 3.3.2 that a smooth space curve defined by the intersection of two surfaces can be projected to a plane curve with at most nodes as singularities. The following is no doubt well known.

Lemma 5 *Let C be a space curve defined by the intersection of two algebraic surfaces. Then we can choose a projection map in NC with respect to the degrees of the polynomials defining the surfaces so that the image of C is a plane curve with at most ordinary nodes as singularities.*

Proof

Recall that a projection is defined as follows: choose a point p off of both the curve C and a plane P. Let q be any point on C. Then the unique line defined by p and q intersects the plane P in exactly one point. The projection maps q to this intersection point.

By [21, pages 132-135] or [12, Theorem IV.3.10], a point will give rise to a bad projection if it lies on a multisecant of the curve (i.e. a secant intersecting the curve in more than two places), a tangent of the curve, or a secant with coplanar tangent lines. In these references, it is shown that the set of bad projection points forms a proper algebraic subset. It can be checked that the polynomials defining this algebraic set are bounded by the degrees of the polynomials defining the surfaces. Thus by [25], and using arguments similar to those that will be given for theorem eith in section four, we can, in NC with respect to the degrees of the polynomials defining the surfaces, choose a point of projection.

One technical note is needed. Both [21] and [12] work over the complex numbers. But a proper algebraic subset of $\mathbf{C^n}$ cannot contain all of the underlying real points. Thus we are indeed guaranteed a good point of projection.

4 Reduction to Bivariate Factorization

All of the previous work depended on the original polynomial being in two variables. In this section we show how to reduce the problem of factoring a multivariable polynomial to the two variable case.

There have been a number of papers giving reductions from multivariate to bivariate factorization. The first appeared in Heintz and Sievking [13], and made use of Bertini's theorem, as will we. This was a randomized irreducibility test that worked for sparse multivariate polynomials. The idea was extended to factorization in [10]. In [16] a reduction was given which is in deterministic polynomial time if the number of variables is fixed, or if the polynomials are dense. [17] later gave a different randomized reduction for the sparse case. These randomized reductions work for polynomials represented as straight-line programs as well as sparse polynomials. An NC reduction for the dense case was given in [15].

For the complex case, we give a new randomized reduction which requires fewer bits per random coefficient $O(\log d)$ than the previous methods $O(d)$ for [17] and $O(d^2)$ for [10]. Thus our reduction also runs in deterministic NC if the number of variables is fixed, or if the polynomials are dense. For sparse polynomials, the reduction is in random NC in the degree d, number of variables n, coefficient size c and number of non-zero terms s. For straight-line program polynomials, the parallel running time is the sum of a polylogarithmic function of measures d, n, c, plus the time to evaluate the polynomial at an integer point.

Given a polynomial $P(x_1, \ldots, x_n)$, we assume that it is square-free. We wish to develop a constructive version of Bertini's theorem, which states that the intersection of an irreducible variety with a generic plane will be an irreducible plane curve. Luckily the proof given in Mumford as Theorem 4.17 in [21] is constructive. Next we describe algebraically the set of intersecting planes that violate Bertini's theorem and then, using Schwartz's lemma [25], show how to choose an intersecting plane satisfying Bertini's theorem.

We use the following (which is Theorem 4.17 in [21]):

Theorem 7 *Let $X \subset \mathbf{P}^n$ (complex projective n-space) be an r-dimensional projective variety and let $M^{n-r-1} \subset \mathbf{P}^n$ be a linear space disjoint from X. Let $p : X \to \mathbf{P}^n$ be the projection from M and let*

$$B = (x \in \mathbf{P}^r \mid \text{p not smooth over x})$$

so that

$$(X - p^{-1}B) \to \mathbf{P}^r - B$$

is a finite-sheeted covering space. Let $l \subset \mathbf{P}^r$ be any line that meets B

transversely. Then

$$p^{-1}(l - l \cap B) \to l - l \cap B$$

is also a connected covering space, hence p^{-1} is an irreducible curve.

In fact by Corollary (4.18) in [21], the above line l can be chosen so that $p^{-1}(l)$ will intersect the variety X transversely. For our case, X will be an irreducible hypersurface $V(P(x_0, \ldots, x_n))$, where $P(x_0, \ldots, x_n)$ is the homogenized version of a polynomial $P(x_1, \ldots, x_n)$. Thus $V(P(x_0, \ldots, x_n))$ is the projective closure of the affine hypersurface $V(P(x_1, \ldots, x_n))$. Then M^{n-r-1} will be for us simply a point off of $V(P(x_0, \ldots, x_n))$.

Following [21], we will find three points that span the plane $p^{-1}(l)$. The first point, p_0 will be the point of projection. Any point off of the hypersurface $V(P(x_0, \ldots, x_n))$ will work.

As in the theorem, let B be the set of points, in the hyperplane $\mathbf{P}^{n-1}$, over which the projection from p_0 is not smooth. In our terminology, B is the set of critical values. We now need to find a line l in the hyperplane $\mathbf{P}^{n-1}$ that is transverse to B. But this is easy. In $\mathbf{P}^{n-1}$, choose a point p_1 off of B and then project B from this point to some $\mathbf{P}^{n-2}$. Let B' be the critical values in $\mathbf{P}^{n-2}$ of this projection map. and let our third point p_2 be a point off of B'. The line l will be defined by the points p_1 and p_2 and the plane $p^{-1}(l)$ will be spanned by p_0, p_1 and p_2. Again, the justification for these choices is in [21].

The following choices must be made. First we must find a point p_0 off of a degree ($d = \deg P$) hypersurface in $\mathbf{P}^n$. Then we must find a point p_1 off of B, which is a degree $d(d-1)$ hypersurface in $\mathbf{P}^{n-1}$. Note that the degree of B is $d(d-1)$ since it is given by the resultant of the polynomial P and some first partial of P. Finally we must find the point p_2 off of the set B', which is a hypersurface of degree $d(d-1)(d(d-1)-1)$ in $\mathbf{P}^{n-2}$. Luckily it is not hard to choose points off of a proper algebraic set. Further, since the set of bad points is a proper algebraic set, we can assume, even though the statement of the above theorem is for projective space, that we are in an affine space $\mathbf{C}^n$ (i.e. we can dehomogenize, since we will only be loosing the points on the hyperplane at infinity, which is a proper algebraic subset of degree one).

Theorem 8 *Let $P(x_1, \ldots, x_n)$ be an irreducible polynomial of degree d. Let E be a finite subset of $\mathbf{C}$. Then the probability that the bivariate polynomial $Q(x, y)$ defining the plane curve, given by the intersection of $V(P)$ by a plane*

spanned by three points in E^n, is reducible is less than $(d^4 - 2d^3 + d^2 + d + 1)/|E|$, where $|E|$ is the cardinality of E.

Proof We make use of Schwartz's lemma [25] that the number of points in the set E^n (E a finite subset of $\mathbf{C}$) that lie in an algebraic set $Z \subset \mathbf{C}^n$ of degree d is at most $d|E|^{n-1}$.

The set of points off of which we want to choose our three points is the union of $V(P)$, B, B' and the hyperplane at infinity, and hence has degree $d + d(d-1) + d(d-1)(d(d-1)-1)$, which is $d^4 - 2d^3 + d^2 + d + 1$. The result follows. □

Corollary 1 *Let $P(x_1, \ldots, x_n)$ be a polynomial of degree d with k irreducible, distinct factors. Let E be a finite subset of $\mathbf{C}$. Then the probability that the bivariate polynomial $Q(x, y)$ defining the plane curve given by the intersection of $V(P)$ by a plane spanned by three points in E^n does not have k factors with corresponding degrees is less than $d^4 - 2d^3 + d^2 + d + 1/|E|$.*

This follows because the points can be chosen exactly as in Theorem (8).

So to achieve a probability of failure less than ϵ, we make sure $|E| > d^4 - 2d^3 + d^2 + d + 1/\epsilon$. Choosing integer values for elements of E therefore requires $(4 \log d + \log \frac{1}{\epsilon})$ bits. For a deterministic algorithm, we take $|E| = d^4 - 2d^3 + d^2 + d$.

Finally, note that by Bertini's theorem almost every reduction to two variables will work. Thus if we do not fix the number of variables, our algorithm will run in random (Monte-Carlo) NC.

5 Factorization Information

We now want to approximate each factor of the polynomial $P(z_1, ..., z_n)$, which is possible due to the recent work of C.A. Neff [22] on approximating the roots of a one variable polynomial with rational coefficients in NC. The arguments used are very straightforward, so we will only sketch the proof.

Let our polynomial $P(z_1, ..., z_n) = \prod P_i(z_1, ..., z_n)$, where each P_i is irreducible of degree d_i. We can assume, after a change of coordinates, that P and the P_i are monic in the variable z_n. There are then $\frac{(d_i+n)!}{(d_i)!(n)!} - 1$ unknown coefficients for each P_i. We will now determine how to approximate these coefficients by solving an associated system of linear equations: $AX = B$,

where A will be an integral invertible matrix, X will be a column vector of coefficients for P_i, and B will be a column vector of algebraic numbers.

Let $a_1, a_2, ..., a_{n-1}$ be integers. Using [22], approximate the roots of the one variable polynomial $P(a_1, a_2, ..., a_{n-1}, z_n)$. Assume for a moment that we can determine which roots are associated to which irreducible factor P_i and can thus approximate the one variable polynomial $P_i(a_1, a_2, ..., a_{n-1}, z_n)$. Then given an integer a_n, we can approximate the algebraic number $b = P(a_1, a_2, ..., a_{n-1}, a_n)$. By choosing $\frac{(d_i+n)!}{(d_i)!(n)!} - 1$ -tuples of integers $(a_1, a_2, ..., a_{n-1}, a_n)$ and treating the coefficients of the P_i as unknowns, we can approximate the coefficients by solving a linear system $AX = B$. Here B is the column vector of the algebraic numbers b from the various $P_i(a_1, a_2, ..., a_{n-1}, a_n)$ and A is the square matrix arising from evaluating all monomials of degree d_i in n variables at the points $(a_1, a_2, ..., a_{n-1}, a_n)$. We must choose our tuple so that the matrix A is invertible, but this is clearly no problem.

There is one difficulty with this method. We do not yet know how to determine which roots of $P(a_1, a_2, ..., a_{n-1}, z_n)$ are associated to which factor P_i. We do know how to do this in the two variable case. For a polynomial $P(z_1, z_2) = \prod P_i(z_1, z_2)$, we can determine which roots of $P(a_1, z_2)$ are associated to which factors P_i, since this is precisely the information that is contained in the connectedness of the earlier constructed curve skeleton, provided that we choose the point a_1 to be on the grid in the z_1 plane, which we can do by enlarging the grid. The general case is now easy. Given two tuples $(a_1, a_2, ..., a_{n-1})$ and $(b_1, ..., b_{n-1})$, intersect $P(z_1, ..., z_n) = 0$ with the plane parallel to the z_n axis containing the points $(a_1, a_2, ..., a_{n-1}, 0)$ and $(b_1, ..., b_{n-1}, 0)$. This reduces the problem of associating roots of P to the two variable case, which we can do. Of course, we cannot intersect $P(z_1, ..., z_n) = 0$ with any plane, but this is not a true difficulty. In the previous section we have an algebraic description of the planes that do not intersect $P(z_1, ..., z_n) = 0$ correctly. Thus we simply must choose our $(n-1)$ tuples so that resulting planes perform properly.

Acknowledgements

We would like to thank Jim Renegar for his comments and discussion of this work. We would also like to thank the referee for many suggestions for improving the exposition of the paper. An earlier version of this work was presented at ISSAC '89, in Portland, Oregon.

References

[1] Borodin, A., von zur Gathen, J. and Hopcroft, J. (1982), "Fast parallel matrix and GCD computations," *Inf. and Contr.*, Vol. 52, pp. 241-256.

[2] Canny, J. (1987), "A New Algebraic Method for Robot Motion Planning and Real Geometry," *28'th Symposium on Foundations of Computer Science*, pp. 39-48.

[3] Canny, J. (1988), "Some Algebraic and Geometric Computations in PSPACE," *20'th Symposium on Theory of Computing*, pp. 460-467.

[4] Chistov, A.L., and Grigoryev, D.Y. (1983), "Subexponential-Time Solving Systems of Algebraic Equations I.," Steklov Institute, LOMI preprint E-9-83.

[5] Ciarlet, P. (1989), *Introduction to numerical linear algebra and optimisation*, Cambridge Texts in Applied Mathematics, Cambridge Universtiy Press.

[6] Davenport, J. and Trager, B. (1981), "Factorization over finitely generated fields," *1981 ACM Symposium on Symbolic Algebraic Computation*, pp. 200-205.

[7] DiCrescenzo, C., and Duval, D., (1984) "Computations on Curves", *Eurosam'84*, LICS 174, pp. 100 - 107.

[8] Duval,D. (1987), *Diverses questions relatives au Calcul Formel Avec Des Nombres Algebriques*, Thèse, L'Université Scientifique, Technologique et Médicale de Grenoble.

[9] Dvornicich , R., and Traverso, C., (1987) "Newton Symmetric Functions and the Arithmetic of Algebraically Closed Fields", in *Proc. AAECC-5*, Springer Lecture Notes Computer Science No. 356., pp. 216-224.

[10] von zur Gathen, J. (1985), "Irreducibility of Multivariate Polynomials," *Journal of Computer System Science*, No. 31, pp. 225-264.

[11] Griffiths, P. and Harris, J. (1978), *Principles of Algebraic Geometry*, John Wiley and Sons

[12] Hartshorne, R. (1977), *Algebraic Geometry*, Springer-Verlag.

[13] Heintz, J. and Sieveking, M. (1981), "Absolute primality of polynomials is decidable in random polynomial time in the number of variables," Proc. 1981 Internat. Conf. Automata, Languages, Prog., *Springer Lec. Notes Comp. Sci.*, Vol. 115, pp. 16-28.

[14] Henrici, P. (1988) *Applied and Computational Complex Analysis*, John Wiley and Sons

[15] Kaltofen, E. (1985), "Fast Parallel Absolute Irreducibility Testing," *J. Symbolic Computation*, Vol. 1, pp. 57-67.

[16] Kaltofen, E. (1985), "Polynomial-Time Reductions from Multivariate to Bi- and Univariate Integral Polynomial Factorization," *SIAM J. Computing*, Vol. 14, pp. 469-489.

[17] Kaltofen, E. (1985), "Effective Hilbert Irreducibility," *Inf. and Contr.*, Vol. 66, No. 3, pp. 123-137.

[18] Kaltofen, E. (1990), "Effective Noether Irreducibility Forms and Applications", *Proc. 23rd Annual Symposium on Theory of Computing*, pp. 54-63.

[19] Kendig, K. (1977) *Elementary Algebraic Geometry*, Springer-Verlag.

[20] Lenstra, A., Lenstra, H. and Lovasz, L. (1982), "Factoring Polynomials with rational coefficients," *Math. Ann.*, Vol. 261, pp. 515-534.

[21] Mumford, D. (1970), *Algebraic Geometry I: Complex Projective Varieties*, Springer-Verlag.

[22] Neff, C. A. (1990), "Specified precision polynomial root isolation is in NC," *Proc. 31st Annual Symposium on Foundations of Computer Science*, pp. 152-162.

[23] Noether, E. (1922), "Ein algebraisches Kriterium fur absolute Irreduzibilitat," *Math. Ann.*, Vol. 85, pp. 26-33.

[24] Pan, V. (1985), "Fast and Efficient Algorithms for Sequential Evaluation of Polynomial Zeroes and of Matrix Polynomials," *26th IEEE Symposium on Foundations of Computer Science.*

[25] Schwartz, J.T. (1980), "Fast Probabilistic Algorithms for Verification of Polynomial Identities," *Jour. ACM*, Vol. 27, No. 4, pp. 701-717.

[26] Shafarevich, I. (1974), *Basic Algebraic Geometry*, Springer Verlag.

[27] Valiant, L. G., Skyum, S., Berkowitz, S., and Rackoff, C., (1983), "Fast Parallel Computation of Polynomials Using Few Processors," *SIAM J. Comp.*, Vol. 12, No. 4, pp. 641-644.