Factors of Low Individual Degree Polynomials

Rafael Oliveira*

Abstract

In [Kal89], Kaltofen proved the remarkable fact that multivariate polynomial factorization can be done efficiently, in randomized polynomial time. Still, more than twenty years after Kaltofen's work, many questions remain unanswered regarding the complexity aspects of polynomial factorization, such as the question of whether factors of polynomials efficiently computed by arithmetic formulas also have small arithmetic formulas, asked in [KSS14], and the question of bounding the depth of the circuits computing the factors of a polynomial.

We are able to answer these questions in the affirmative for the interesting class of polynomials of bounded individual degrees, which contains polynomials such as the determinant and the permanent. We show that if $P(x_1, \ldots, x_n)$ is a polynomial with individual degrees bounded by r that can be computed by a formula of size s and depth d, then any factor $f(x_1, \ldots, x_n)$ of $P(x_1, \ldots, x_n)$ can be computed by a formula of size $\mathsf{poly}((rn)^r, s)$ and depth $d+5$. This partially answers the question above posed in [KSS14], who asked if this result holds without the dependence on r. Our work generalizes the main factorization theorem from Dvir et al. [DSY09], who proved it for the special case when the factors are of the form $f(x_1, \ldots, x_n) \equiv x_n - g(x_1, \ldots, x_{n-1})$. Along the way, we introduce several new technical ideas that could be of independent interest when studying arithmetic circuits (or formulas).

*Department of Computer Science, Princeton University. Research supported by NSF grant CCF-1217416 and by the Sloan fellowship. Email: `rmo@cs.princeton.edu`.

Contents

1 Introduction

Let $f(x_1,\ldots,x_n) \in \mathbb{F}[x_1,\ldots,x_n]$ be a multivariate polynomial over a field $\mathbb{F}$. The individual degree of f with respect to variable x_i, denoted by $\deg_{x_i}(f)$, is the largest power of x_i appearing in a monomial of f. Many interesting polynomials have bounded individual degree, such as the Permanent and Determinant polynomials. Moreover, the class of polynomials of bounded individual degree is closed under factorization, since if a polynomial $f(x_1,\ldots,x_n)$ has individual degrees bounded by r, so will its factors. In this work, we study the problem of formula (circuit) factorization of polynomials of low individual degree.

One of the basic operations on polynomials is factorization. This problem can be phrased as follows: given a polynomial $P(x_1,\ldots,x_n)$, decide whether $P(x_1,\ldots,x_n)$ is irreducible, or if not, output one of its factors, which we denote by $f(x_1,\ldots,x_n)$. From the computational perspective, we will usually be given a device computing the polynomial P, and we will be asked to output a similar device computing f. In the field of arithmetic complexity, the most natural device computing polynomials is an arithmetic circuit or a formula (see Definition 1.1 below). Therefore, we will assume that we are given P as an arithmetic circuit (formula) and output one of its factors in the same representation. We now give the definition of an arithmetic circuit/formula:

Definition 1.1. *An* arithmetic circuit Γ *is a directed acyclic labeled graph in which the vertices are called 'gates'. The gates of* Γ *with in-degree* 0 *are called* inputs *and are labeled by either a variable from* $\{x_1,\ldots,x_n\}$ *or by field element from* $\mathbb{F}$. *Every other gate of* Γ *is labeled by either '*$\times$*' or '*$+$*' and has in-degree 2. (If we talk about bounded depth circuits/formulas, then we remove the restriction on the in-degree.) There is one gate with out-degree 0, which we call the* output gate. *Each gate in* Γ *computes a polynomial in* $\mathbb{F}[x_1,\ldots,x_n]$ *in the natural way. An arithmetic circuit is called a* formula *if its underlying graph is a tree. The* size *of a circuit (formula)* Γ, *written* $|\Gamma|$, *is given by the number of edges in the circuit (formula) and the* depth *of* Γ, *written* $\mathrm{depth}(\Gamma)$*, is defined as the length of the longest directed path in the graph of* Γ.

Polynomial factorization is one of the cornerstone problems in modern computer algebra, and as such has been the focus of intensive research. The past three decades have seen major advances on the development of efficient algorithms for polynomial factorization, pioneered by the works of Lenstra et al. and Kaltofen [LLL82, Kal85, Kal89, Kal03]. In addition to the general problem, polynomial factorization has also been studied in many other important (and more restricted) representations. For instance, in the sparse representation, where the input polynomial is given as a list of its coefficients and monomials, the works of Lenstra, Kaltofen and von zur Gathen [LJ99, GK85] give efficient algorithms for sparse factorization in the univariate and in the multivariate cases. For a more complete survey on polynomial factorization we refer the reader to the survey [Kal03] and to the book [GG99].

In the seminal work of Kaltofen [Kal89], it is proved that if $P(x_1,\ldots,x_n)$ of total degree D can be computed by an arithmetic circuit of size s, then any of its factors have arithmetic circuits of size $\mathsf{poly}(n,s,D)$. Moreover, Kaltofen gives a randomized algorithm that with high probability outputs such a factor in polynomial time. This result, besides settling an important complexity theoretic question, has since then had a great impact in many areas of computer science, such as coding theory [Sud97, GS06], derandomization [KI04] and cryptography [CR88]. However, many interesting questions on the complexity of arithmetic circuits or formulas under factorization remain unanswered. In particular, we study the following two questions, where the first one was asked in the work of Kopparty et al. [KSS14], while the second question was stated as an open problem in the survey [SY10, Open Problem 19]:

1. If $P(x_1, \ldots, x_n)$ of total degree D is computed by an arithmetic formula of size s, is it true that any of its factors will also have formulas of size $\mathsf{poly}(n, s, D)$?

2. If $P(x_1, \ldots, x_n, y)$ can be computed by a circuit of size s and depth d, can its factors be computed by a circuit of size $\mathsf{poly}(s)$ and depth $O(d)$?

In this work, we answer both of these questions in the affirmative, in the case where the input polynomial P has bounded individual degrees. In particular, we show:

Theorem 1. *Let $P(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n] \setminus \{0\}$ be such that $\deg_{x_i}(P) \le r,\ 1 \le i \le n$, and let $f(x_1, \ldots, x_n) \in \mathbb{F}[x_1, \ldots, x_n]$ be a factor of P, where $\mathbb{F}$ is a field of characteristic zero. If there exists a formula (circuit) of size s and depth d computing P, then there exists a formula (circuit) of depth $d + 5$ and size $\mathsf{poly}((nr)^r, s)$ that computes $f(x_1, \ldots, x_n)$. Moreover, if we require the in-degree of each gate to be 2, then the size remains the same and the depth becomes $d + O(r \log(nr))$.*

Remark 1.2. *In addition to the structural result above, our proof also yields a randomized algorithm for finding all factors of $P(x_1, \ldots, x_n)$ in time $\mathsf{poly}((nr)^r, s)$. The output of our algorithm is precisely the formulas (circuits) described in Theorem 1.*

Notice that our theorem has no restriction on the individual degrees of the polynomials computed by the intermediate gates of the circuit (that is, we have no syntactic restrictions). We only care about the individual degrees of the output polynomial, which we regard as bounded by a constant, denoted by r, in the theorem above.

Theorem 1 provides a direct answer to the second question posed above in the case where P has bounded individual degrees (that is, r is a constant). The connection between Theorem 1 and the first question comes from the fact that one can always balance formulas to have logarithmic depth. More precisely, suppose that we are given a formula Φ (with in-degree bounded by 2) of size $s = \mathsf{poly}(n)$ computing P. By Theorem 2.7 in [SY10], we can assume that Φ is of size $\mathsf{poly}(s)$ and $\mathrm{depth}(\Phi) = O(\log s)$. Hence, Theorem 1 implies that there exists a formula Ψ, with in-degree bounded by 2, of depth $\mathrm{depth}(\Psi) = \mathrm{depth}(\Phi) + O(r \log(sn)) = O(\log s)$ and size $\mathsf{poly}((nr)^r, s) = \mathsf{poly}(s)$ computing any factor $f(x_1, \ldots, x_n)$ of P. This provides an affirmative answer to the first question.

Before giving an overview of the proof of Theorem 1, we give some background on related work on factorization in general and in bounded depth circuits.

The problem of factoring in bounded depth was studied previously in [DSY09], who showed that if $P(x_1, \ldots, x_n)$ has a depth d circuit of size s and $\deg_{x_n}(P) \le r$, then its factors of the form $x_n - \phi(x_1, \ldots, x_{n-1})$ have depth $d + 3$ circuits of size $\mathsf{poly}(n^r, s)$. This result was used to extend the hardness-randomness tradeoffs of [KI04] to the bounded depth model. Our main theorem generalizes their result to any factor of P, provided that P has bounded individual degrees.

Shpilka and Volkovich in [SV10] initiated the study of factorization of multilinear polynomials, which are the most basic case of polynomials of bounded individual degrees. They relate the problem of deterministically factoring multilinear polynomials to the problem of performing deterministic Polynomial Identity Testing (PIT). In their paper, they prove that these two problems are roughly equivalent in the multilinear setting for most restricted multilinear circuit classes that have been studied. Since the problem of performing deterministic PIT seems to be hard, even for the class of multilinear formulas, this shed some light on the difficulty of obtaining deterministic factorization even for this model. This equivalence between deterministic PIT and deterministic polynomial factorization was later generalized by Kopparty et al. in [KSS14] to polynomials (of polynomial degree) computed by general circuits. Since we prove here that, for polynomials of bounded individual degrees computed by circuits of small depth, their factors can also be computed by circuits of small depth, one could hope for similar connections between PIT for restricted classes of circuits – say of bounded depth and low individual degrees – and factorization of polynomials in such classes.

1.1 Proof Overview

In this section, we give an overview of the proof of the main theorem. For simplicity of exposition, we will only refer to arithmetic circuits in this overview, but our results hold true for formulas as well, as the proofs and the statements in the later sections show. We begin with the simplest example, where our polynomial is *monic* with respect to one of the variables and also *factors completely* with respect to this variable, that is:

$$P(x_1, \ldots, x_n, y) \equiv \prod_{i=1}^{r} (y - g_i(x_1, \ldots, x_n)),$$

where each polynomial $g_i(x_1, \ldots, x_n)$ has a nonzero constant term μ_i and $\mu_i \neq \mu_j$ for $i \neq j$[1]. For this example, we refer to the polynomials $g_i(x_1, \ldots, x_n)$ as the roots of $P(x_1, \ldots, x_n, y)$, since $P(x_1, \ldots, x_n, g_i(x_1, \ldots, x_n)) \equiv 0$. Now, the big question is: how can we find the roots of $P(x_1, \ldots, x_n, y)$ with respect to variable y?

In this case we are in the framework of [DSY09], since

$$P(0, \ldots, 0, y) \equiv \prod_{i=1}^{r} (y - \mu_i)$$

and the roots μ_i are distinct. The main idea in Dvir et al. [DSY09] is to build a circuit for the roots $g_i(x_1, \ldots, x_n)$ by iteratively finding the homogeneous parts of $g_i(x_1, \ldots, x_n)$ one at a time. That is, they show how to construct polynomials $q_{i,t}(x_1, \ldots, x_n)$ which agree with the roots $g_i(x_1, \ldots, x_n)$ on all homogeneous parts up to degree t. Moreover, they show that if P

[1] As Section 3 shows, we can guarantee distinct roots in $P(0, \ldots, 0, y)$ by using a random shift of the variables $(x_1, \ldots, x_n)$.

is computed by a circuit Γ of size s and depth d, then there exists a circuit of size $\mathsf{poly}(t^r, s)$ and depth $d+2$ computing $q_{i,t}(x_1, \ldots, x_n)$.

The polynomial $q_{i,t}(x_1, \ldots, x_n)$, which approximates the root $g_i(x_1, \ldots, x_n)$, satisfies the following property: $P(x_1, \ldots, x_n, q_{i,t}(x_1, \ldots, x_n))$ is a polynomial which only has monomials of degree larger than t. This motivates us to make the following definition:

Definition 1.3 (Approximate Root)**.** *Let $P(x_1, \ldots, x_n, y)$ be a polynomial in $\mathbb{F}[x_1, \ldots, x_n, y]$. We say that $q(x_1, \ldots, x_n)$ is a* root *of P* up to degree t *if all the homogeneous parts up to degree t of the polynomial $P(x_1, \ldots, x_n, q(x_1, \ldots, x_n))$ are zero. That is, $P(x_1, \ldots, x_n, q(x_1, \ldots, x_n))$ only has monomials of degree larger than t.*

With this definition in hand, we can show that finding an approximate root $q_{i,t}(x_1, \ldots, x_n)$ of $P(x_1, \ldots, x_n, y)$ corresponds to finding the homogeneous parts of an actual root $g_i(x_1, \ldots, x_n)$. Here is a basic version of what is done in Dvir et al. [DSY09], which shows that an approximate root $q_{i,t}(x_1, \ldots, x_n)$ in actuality approximates a root $g_i(x_1, \ldots, x_n)$ of $P(x_1, \ldots, x_n, y)$.

For each μ_i and $t \geq 1$, we can find polynomials $q_{i,t}(x_1, \ldots, x_n)$ such that $q_{i,t}(0, \ldots, 0) = \mu_i$ and the polynomial $P(x_1, \ldots, x_n, q_{i,t}(x_1, \ldots, x_n))$ only has terms of degree larger than t. Since

$$P(x_1, \ldots, x_n, q_{i,t}(x_1, \ldots, x_n)) \equiv \prod_{j=1}^{r} (q_{i,t}(x_1, \ldots, x_n) - g_j(x_1, \ldots, x_n)),$$

the minimum degree terms of $P(x_1, \ldots, x_n, q_{i,t}(x_1, \ldots, x_n))$ must come from the product of the minimum degree terms of each of the polynomials $q_{i,t}(x_1, \ldots, x_n) - g_j(x_1, \ldots, x_n)$. Notice that for each $j \neq i$ the constant term of each polynomial $q_{i,t}(x_1, \ldots, x_n) - g_j(x_1, \ldots, x_n)$ is equal to $\mu_i - \mu_j$, which is nonzero. Therefore, the minimum degree terms of $P(x_1, \ldots, x_n, q_{i,t}(x_1, \ldots, x_n))$ must come from the minimum degree terms of the polynomial $q_{i,t}(x_1, \ldots, x_n) - g_i(x_1, \ldots, x_n)$. Because $P(x_1, \ldots, x_n, q_{i,t}(x_1, \ldots, x_n))$ only has terms of degree larger than t, the same must happen to the polynomial $q_{i,t}(x_1, \ldots, x_n) - g_i(x_1, \ldots, x_n)$. This implies that $q_{i,t}(x_1, \ldots, x_n)$ approximates the actual root $g_i(x_1, \ldots, x_n)$ of P up to degree t. Hence, if we pick t larger than the total degree of g_i, we obtain that the lower degree terms of $q_{i,t}$ correspond to the root g_i, and therefore we can recover this root g_i (and use them to factor P).

There are two main issues with this approach that we need to overcome, if we are to generalize it. The first issue is that P may *not factor into linear factors* in y, that is, polynomials of the form $y - g_i(x_1, \ldots, x_n)$. The second one is that P need *not be monic* in y, in which case we will still need to recover its leading coefficient – which is a polynomial in $\mathbb{F}[x_1, \ldots, x_n]$. We will deal with these two issues in the following two subsections.

1.1.1 Approximating Non-Linear Roots

To deal with the first issue, let us study a toy example: assume that P is monic in y with $\deg_y(P) = r$, that is,

$$P(x_1, \ldots, x_n, y) \equiv y^r + \sum_{i=0}^{r-1} P_i(x_1, \ldots, x_n) y^i,$$

but P does not factor into linear factors in y. Let $f(x_1, \ldots, x_n, y)$ be one of its factors, of degree k in y. Since P is monic in y, we know that f must also be monic in y. Note that if we work over the algebraic closure of $\mathbb{F}(x_1, \ldots, x_n)$ (that is, the field $\overline{\mathbb{F}(x_1, \ldots, x_n)}$), we can factor P (and f) into linear factors in y. In this work, we will not describe what the algebraic closure of $\mathbb{F}[x_1, \ldots, x_n]$ is, since it is a very complex field, and it is not needed in our proof. We only mention $\overline{\mathbb{F}(x_1, \ldots, x_n)}$ here to give us some intuition on how to generalize the root finding approach described above. For simplicity, simply think of elements of the closure as "functions" over the variables $x_1, \ldots, x_n$. Since f divides P, if

$$P(x_1, \ldots, x_n, y) \equiv \prod_{i=1}^{r} (y - \varphi_i(x_1, \ldots, x_n)),$$

then there will be indices (say i from 1 to k) such that

$$f(x_1, \ldots, x_n, y) \equiv \prod_{i=1}^{k} (y - \varphi_i(x_1, \ldots, x_n)).$$

However, it is worth noting that these linear factors will not be polynomials! Nevertheless, the fact that they share some roots in the closure of $\mathbb{F}[x_1, \ldots, x_n]$ gives us a hint on what to do next. To overcome this problem, we will (in Lemma 5.1 and Corollary 5.2) approximate these functions φ_i by polynomials $g_{i,t}$, in a way that the polynomial

$$g_t(x_1, \ldots, x_n, y) \equiv \prod_{i=1}^{k} (y - g_{i,t}(x_1, \ldots, x_n))$$

agrees with f on the terms of order smaller than t. Therefore, for large enough t, we will have that the lower order terms of $g_t(x_1, \ldots, x_n, y)$ will correspond to the polynomial f, which we can then obtain by interpolation (Lemma 2.3). We can think of each polynomial $g_{i,t}$ as the Taylor expansion of φ_i up to degree t.

The way we obtain these approximations to the roots (the polynomials $g_{i,t}$) is by a procedure similar in nature to Hensel lifting. Suppose that $\varphi_i(0, \ldots, 0) = \mu_i$ for $1 \leq i \leq k$, and moreover, suppose that $\mu_i \neq \mu_j$ for $i \neq j$. From each valuation μ_i, we will construct a family of polynomials $g_{i,t}$ of degree t, such that $g_{i,t}(x_1, \ldots, x_n)$ is a root of f up to degree t. Now, the question is: how can we construct this family of polynomials if we do not have access to f? The answer to this question lies on the fact that each root $y - \varphi_i$ of f is also a root of P, and therefore we can access the valuations of φ_i's through the circuit computing P. Hence, we will use the fact that the φ_i's are also roots of P in order to find the polynomials g_t that approximate f (Lemma 6.1).

We are now ready to tackle the second issue, which happens when our input polynomial $P(x_1, \ldots, x_n, y)$ is not monic.

1.1.2 Recovering the Leading Coefficient

To overcome the second main issue, that the polynomial P may not be monic, let us define

$$f(x_1, \ldots, x_n, y) \equiv \sum_{i=0}^{k} f_i(x_1, \ldots, x_n) y^i \quad \text{and} \quad P(x_1, \ldots, x_n, y) \equiv \sum_{i=0}^{r} P_i(x_1, \ldots, x_n) y^i,$$

where $f_k(x_1, \ldots, x_n) \not\equiv 0$ and $P_r(x_1, \ldots, x_n) \not\equiv 0$. If f divides P, then it must be the case that the leading coefficient f_k of f divides the leading coefficient P_r of P. Hence, a possible solution to this second issue would be to find, by some kind of induction, a small circuit for f_k based on the circuit for P_r that we obtain from P. Then, we could generalize the factoring result for monic polynomials to the case where the factors are rational functions of the form

$$\frac{f(x_1, \ldots, x_n, y)}{f_k(x_1, \ldots, x_n)} \equiv y^k + \sum_{i=0}^{k-1} \frac{f_i(x_1, \ldots, x_n)}{f_k(x_1, \ldots, x_n)} y^i.$$

With these two results, we could multiply the formulas computing f_k and $\dfrac{f}{f_k}$ to obtain our factor f.

More precisely, if we could find, by induction on the number of variables, a small formula Φ_k for f_k based on the formula Γ_r for P_r that we obtain from P via interpolation (Lemma 2.4), and if we could find a small formula Υ for the rational function $\dfrac{f}{f_k}$ based on the formula Γ computing P (Lemma 6.1), then the formula given by $\Upsilon \times \Phi_k$ would compute the polynomial f, as we wanted.

One problem with this approach is that, even if we can generalize the monic factoring result to monic rational functions as above, as far as we know, the best bound on the size of the circuit Γ_r computing P_r is given by $3r \cdot s$ (see Lemma 2.4). Therefore, if we define $T(n, s)$ as the maximum size of a factor of a polynomial in n variables computed by a circuit of size s, the induction given by the procedure above would give us the following bounds on the size:

$$T(n+1, s) \leq T(n, 3r \cdot s) + \mathsf{poly}((nr)^r, s).$$

The reason for this bound is the following: $P(x_1, \ldots, x_n, y)$ has $n+1$ variables and is computed by Γ, which has size s. Hence, the maximum size of a factor f is by definition $T(n+1, s)$. Since f_k divides the leading coefficient P_r, which is computed by Γ_r of size $3rs$ and has n variables, the bound we have on the size of Φ_k is given by $T(n, 3rs)$, because now the input polynomial is P_r. Assuming that the size of f/f_k can be bounded by $((nr)^r \cdot s)^\alpha$, for some constant α (which we can by Lemma 6.1), we obtain the additive factor $\mathsf{poly}((nr)^r, s)$. Since the circuit for f is given by $\Upsilon \times \Phi_k$, we need to add the bounds on the sizes for Φ_k and Υ. However, when we solve this equation, we obtain that

$$T(n+1, s) \leq T(1, (3r)^n \cdot s) + \mathsf{poly}((nr)^r, (3r)^n \cdot s)$$

which is exponential in n, the number of variables! Therefore, this approach, as it is, cannot work.

The main problem in the recursion above is that the bound on the circuit size of the leading coefficient, if we only use Lemma 2.4, keeps getting worse as we reduce the number of variables – it will become $(3r)^\ell \cdot s$ if we get rid of ℓ variables. One way to get around this issue is to find a coefficient $P_i(x_1, \ldots, x_n)$ which has smaller circuit size than $P(x_1, \ldots, x_n, y)$. This would avoid the exponential blow-up in the recursion above. Notice that $P_0(x_1, \ldots, x_n)$ is a very good coefficient in this regard, as $P_0(x_1, \ldots, x_n) = P(x_1, \ldots, x_n, 0)$ and therefore $P_0(x_1, \ldots, x_n)$ has smaller circuit size than $P(x_1, \ldots, x_n, y)$. However, $P_0(x_1, \ldots, x_n)$ is not the leading coefficient of $P(x_1, \ldots, x_n, y)$. The question now becomes: can we make $P_0(x_1, \ldots, x_n)$ the leading coefficient, while still preserving some divisibility properties?

To address this issue, we define the *reversal* of a polynomial with respect to a specific variable and we study its properties with regards to divisibility. If

$$P(x_1, \ldots, x_n, y) \equiv \sum_{i=0}^{r} P_i(x_1, \ldots, x_n) y^i$$

is a polynomial, with $P_r(x_1, \ldots, x_n) \cdot P_0(x_1, \ldots, x_n) \not\equiv 0$, we define its reversal with respect to y as the polynomial

$$\tilde{P}(x_1, \ldots, x_n, y) \equiv \sum_{i=0}^{r} P_i(x_1, \ldots, x_n) y^{r-i}.$$

That is, $\tilde{P}$ is obtained from the polynomial P by "reversing" the coefficients $P_i(x_1, \ldots, x_n)$. It is easy to see that f divides P iff $\tilde{f}$ divides $\tilde{P}$. With this fact in mind, notice that we have transformed the leading coefficient of our problem from $P_r(x_1, \ldots, x_n)$ to $P_0(x_1, \ldots, x_n)$. This has the advantage that now, the leading coefficient of our input polynomial can be computed by the circuit $\Gamma|_{y=0}$ (that is, the circuit obtained from Γ by setting $y = 0$), which has size $\leq s$. This now allows us to recurse into the division of f_0 by P_0 (the new leading coefficients after the reversal) without paying the multiplicative cost on the size of the circuit. Hence with this idea we avoid paying the exponential blowup on the circuit size! On the coin side, notice that the size of the circuit computing the polynomial $\tilde{P}$ is bounded by $8r^2 \cdot s$, according to Lemma 2.8. But this blow up does not hurt us, since the reversal is not cumulative.

More precisely, we now have the following recursion: since we want to bound the size of a factor of P, computed by a circuit Γ of size s and on $n+1$ variables, the bound is by definition $T(n+1, s)$. Now, if we can find a circuit computing f/f_0 from the circuit $\widetilde{\Gamma}$ computing $\tilde{P}$ of size bounded by $((nr)^r \cdot |\widetilde{\Gamma}|)^\alpha \leq ((nr)^r \cdot 8r^2 s)^\alpha$, for some constant α (which we can by Lemma 6.1), we are only left with the problem of finding a small circuit for f_0, which divides P_0, which in turn can be computed by a circuit of size bounded by s in n variables. The bound for a circuit for f_0 is given in this case by $T(n, s)$, by definition of the function T. Therefore, our recursion becomes

$$T(n+1, s) \leq T(n, s) + ((nr)^r \cdot 8r^2 \cdot s)^\alpha$$

which implies that

$$T(n, s) \leq n \cdot ((nr)^r \cdot 8r^2 \cdot s)^\alpha = \mathsf{poly}((nr)^r, s),$$

as we wanted!

The idea of the reversal of a polynomial is similar to the definition of *reversal* of a univariate polynomial given in [GG99, §9.1]. This notion of reversal is used there to perform division with remainder for univariate polynomials by using Newton iteration.

Now, we only need to generalize the root approximation technique to the case where $P(x_1,\ldots,x_n,y)$ is monic in y with coefficients being rational functions in $(x_1,\ldots,x_n)$. This is what we do next.

1.1.3 Root Approximation for Rational Functions

To generalize the monic factoring result to the case when f is monic in y with rational coefficients, we introduce the idea of an approximation polynomial of a rational function (see Section 4), and we use this approximation polynomial in Lemma 6.1 (instead of the rational function) as the "factor" of the input polynomial. If f is a rational function of the form

$$f(x_1,\ldots,x_n,y)\equiv\frac{1}{1-g(x_1,\ldots,x_n)}\cdot\sum_{i=0}^{k}f_i(x_1,\ldots,x_n)y^i,$$

where $g(x_1,\ldots,x_n)$ and $f_i(x_1,\ldots,x_n)$ are polynomials in $\mathbb{F}[x_1,\ldots,x_n]$ such that $g(0,\ldots,0)=0$, we define its approximation polynomial (to degree m) as the following polynomial

$$\psi_{f,m}(x_1,\ldots,x_n,y)\equiv(1+g+g^2+\ldots+g^m)\cdot\sum_{i=0}^{k}f_iy^i,$$

where $g\equiv g(x_1,\ldots,x_n)$ and $f_i\equiv f_i(x_1,\ldots,x_n)$. This polynomial "approximates" the rational function $f(x_1,\ldots,x_n,y)$ in the sense that, for large enough m, the polynomial obtained by $\psi_{f,m}(x_1,\ldots,x_n,y)\cdot(1-g(x_1,\ldots,x_n))$ is equal to $f(x_1,\ldots,x_n)\cdot(1-g(x_1,\ldots,x_n))$, up to high order terms (see Observation 4.3), which we can get rid of by interpolation (Lemma 2.3). By adapting the approach in [DSY09] to work with approximation polynomials, we can find all the "roots" of the approximation polynomials, and after that combine this approximation polynomial with the circuit obtained to compute the leading term.

After we take care of finding the leading coefficient $f_0(x_1,\ldots,x_n)$ (of the reversed polynomial $\tilde{f}(x_1,\ldots,x_n,y)$), and after recovering the approximation polynomial $\psi_{f,m}$ (see Lemma 6.1), we can multiply it by f_0 to obtain the factor f (up to high order tems) which, after interpolation, becomes our desired factor (see Theorem 6.3).

1.1.4 Summary of Main Ideas

We conclude this proof outline with a basic roadmap of the main ideas involved in this work:

1. (Preprocessing) Given a circuit Γ for our polynomial $P(x_1,\ldots,x_n,y)$, we find a circuit $\tilde{\Gamma}$ computing the reversal polynomial $\tilde{P}(x_1,\ldots,x_n,y)$. (Lemma 2.8)

2. (Approximating Roots) We use the circuit $\tilde{\Gamma}$ to find small circuits $\Phi_{i,t}$ for each approximate root of $\tilde{P}$ up to degree t. (Section 5)

3. (Reconstructing factor from approximate roots) Since $\tilde{f}$, divides $\tilde{P}$ (Lemma 2.9), any approximate root of $\tilde{f}$ will also be an approximate root of $\tilde{P}$. By combining the circuits $\Phi_{i,t}$ computing the approximate roots of $\tilde{f}(x_1, \ldots, x_n, y)$, find circuit Ψ computing the approximation polynomial (see Section 4) of the monic rational function $\dfrac{\tilde{f}(x_1, \ldots, x_n, y)}{f_0(x_1, \ldots, x_n)}$. (Lemma 6.1)

4. (Reconstructing leading term by induction) By induction, obtain the circuit Λ_0 computing $f_0(x_1, \ldots, x_n)$, through the circuit $\Gamma|_{y=0}$ computing $P_0(x_1, \ldots, x_n) \equiv P(x_1, \ldots, x_n, 0)$.

5. (Obtaining approximation to the factor) We then prove that the lower order terms of the circuit $\Phi = \Lambda_0 \times \Psi$ compute the polynomial $\tilde{f}$. (Theorem 6.3)

6. (Clean-up stage) By interpolation (Lemma 2.3) and by the Reversal Lemma (Lemma 2.8), obtain the lower order terms from Φ computing f.

1.2 Organization

The rest of the paper is organized as follows: in Section 2 we set up notations, go over some useful background and discuss the concept of reversal. In Section 3 we introduce the concept of properly splitting variable restrictions. In Section 4, we formally introduce the concepts of standard forms and approximation polynomials. In Section 5, we adapt the approach of [DSY09] to find small formulas for the roots of $P(x_1, \ldots, x_n, y)$. In Section 6 we prove our main technical lemma and theorem. In Section 7, we conclude and propose some open problems.

2 Preliminaries

In this section, we establish the notation that will be used throughout the paper and some technical background that we will need to develop the proof of our main theorem. We defer the proofs of the lemmas in this section to the appendix (Section A).

2.1 Notations

From this point on, we will use boldface for vectors, and regular font for scalars. Thus, we will denote the vector $(x_1, \ldots, x_n)$ by $\mathbf{x}$. If we want to multiply the vector $\mathbf{x}$ by a scalar z we will denote this product by $z\mathbf{x}$.

We will denote our base field by $\mathbb{F}$, and we will assume that $\mathbb{F}$ has characteristic zero and that it is algebraically closed. The results in this paper also hold for non-closed fields of characteristic polynomial in the bound on the individual degree, if we allow ourselves to use coefficients from field extensions. More precisely, if we work with polynomials having individual degree at most r, then our result works over all fields of characteristic $O(r^2)$[2]. The assumptions just made are for clarity of exposition.

[2]The polynomial dependency on the bound of the individual degrees is necessary because we need to maintain nonzeroness of some resultants. Since resultants of polynomials of individual degree bounded by r have individual

Let $\mathbb{N}_0$ be the set of natural numbers including zero, that is, $\mathbb{N}_0 = \{0, 1, 2, 3, \ldots\}$. If $\mathbf{e} \in \mathbb{N}_0^n$ is a vector of natural numbers and $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of formal variables, we define $\mathbf{x}^{\mathbf{e}} = \prod_{i=1}^{n} x_i^{e_i}$. That is, $\mathbf{x}^{\mathbf{e}}$ is the monomial corresponding to the product of the variables $\prod_{i=1}^{n} x_i^{e_i}$, where each variable is raised to the proper power.

We will denote $\mathbb{F}(\mathbf{x})[y]$ as the set of polynomials in the variable y whose coefficients are rational functions on the variables $\mathbf{x}$. That is, $f(\mathbf{x}, y) \in \mathbb{F}(\mathbf{x})[y]$ iff it can be expressed in the form $f(\mathbf{x}, y) \equiv \sum_{i=0}^{k} \frac{f_i(\mathbf{x})}{g_i(\mathbf{x})} y^i$, with $f_i(\mathbf{x}), g_i(\mathbf{x}) \in \mathbb{F}[\mathbf{x}], 0 \leq i \leq k$.

When working with a polynomial in $\mathbb{F}[\mathbf{x}, y]$, we might be interested in looking at the homogeneous parts of a polynomial with respect to certain variables only. This will be particularly useful when lifting the "roots" of a polynomial $f(\mathbf{x}, y)$ of the form $y - q(\mathbf{x})$ in order to obtain a circuit computing $f(\mathbf{x}, y)$. To this end, we introduce the following definition.

Definition 2.1 (Partial Homogeneous Parts). *Let $P(\mathbf{x}, y) \equiv \sum_{\mathbf{d}} \alpha_{\mathbf{d}}(y) \cdot \mathbf{x}^{\mathbf{d}}$ be a polynomial in $\mathbb{F}[\mathbf{x}, y]$, where each $\alpha_{\mathbf{d}}(y) \in \mathbb{F}[y]$. For each $m \in \mathbb{N}_0$, we define $H_m^{\mathbf{x}}[P]$ as the polynomial formed by the homogeneous parts of degree m of $P(\mathbf{x}, y)$, when seen as a polynomial in $\mathbb{F}[y][\mathbf{x}]$, that is, when considered as a polynomial on the variables $\mathbf{x}$, and regarding y as a constant. More explicitly, $H_m^{\mathbf{x}}[P]$ is equal to the sum of all monomials of P that have degree m in $x_1, \ldots, x_n$, without any restrictions on the degree of y. We also define $H_{\leq m}^{\mathbf{x}}[P] \equiv \sum_{i=0}^{m} H_i^{\mathbf{x}}[P]$.*

For example, if $P(\mathbf{x}, y) \equiv (x_1x_3x_4 - x_2^3 + x_1x_2)y^2 + (x_1^2x_3 - x_4)y + x_2^2x_3 - x_1x_4$, we have that $H_3^{\mathbf{x}}[P(\mathbf{x}, y)] \equiv (x_1x_3x_4 - x_2^3)y^2 + x_1^2x_3y + x_2^2x_3$.

Notice that if $P(\mathbf{x}, y) \equiv \sum_{i=0}^{r} P_i(\mathbf{x})y^i$, then the partial homogeneous parts satisfy the following property:

$$H_m^{\mathbf{x}}[P(\mathbf{x}, y)] \equiv \sum_{i=0}^{r} H_m^{\mathbf{x}}[P_i(\mathbf{x})] \cdot y^i.$$

Therefore, this definition of partial homogeneous parts agrees with the definition of homogeneous parts if $P(\mathbf{x}, y)$ does not depend on variable y.

When talking about partial homogeneous parts of a polynomial, it is useful to have a notion of minimum degree with respect to some variables.

Definition 2.2 (Minimum Degree). *Let $f(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be a polynomial. We define* $\operatorname{mindeg}_{\mathbf{x}}(f(\mathbf{x}, y))$ *to be the minimum degree of polynomial $f(\mathbf{x}, y)$ on the variables $\mathbf{x}$. In other words, we have* $\operatorname{mindeg}_{\mathbf{x}}(f(\mathbf{x}, y)) = \min_{\ell} (H_\ell^{\mathbf{x}}[f] \not\equiv 0)$ *. For instance, if $f(\mathbf{x}, y) = x_1x_2x_3y - x_1^2x_2^2 + x_3^5$, we have that* $\operatorname{mindeg}_{\mathbf{x}}(f) = 3$.

degree bounded by $O(r^2)$, we need this bound on the characteristic to ensure nonzeroness. We need field extensions because we are performing interpolation of polynomials of high degree.

2.2 Basic Operations on Circuits and Formulas

We begin with the following standard lemma on obtaining the homogeneous components of a polynomial. The version below is from [DSY09].

Lemma 2.3 (Homogeneous Components Through interpolation)**.** *Let $P(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a polynomial with degree $\deg(P) = m$ such that P can be computed by a formula (circuit) Γ of depth d. Then, there exists a formula (circuit) Δ with $m+1$ outputs, of size $|\Delta| \leq 9m^2 \cdot |\Gamma|$ and depth $\mathrm{depth}(\Delta) \leq \mathrm{depth}(\Gamma) + 1$ that computes $H_0^{\mathbf{x}}[P], \ldots, H_m^{\mathbf{x}}[P]$. Moreover, if the topmost gate in the formula (circuit) for $P(\mathbf{x})$ is an addition gate, then we have $\mathrm{depth}(\Delta) = \mathrm{depth}(\Gamma) = d$.*

The next lemma shows us how to obtain the coefficients of a polynomial through interpolation.

Lemma 2.4 (Interpolation)**.** *Let $P(\mathbf{x}, y) \equiv \sum_{i=0}^{r} y^i P_i(\mathbf{x})$ be a polynomial computed by a formula (circuit) Γ. Then for each $i \in \{0, 1, \ldots, r\}$, there exists a formula (circuit) Φ_i such that $|\Phi_i| \leq 3r \cdot |\Gamma|$ and Φ_i computes the polynomial $P_i(\mathbf{x})$.*

As a corollary, we obtain the following lemma:

Lemma 2.5. *Let $P(\mathbf{x}, y) \equiv \sum_{i=0}^{r} y^i P_i(\mathbf{x})$ be a polynomial computed by a formula (circuit) Γ with output gate being an addition gate. Then for each $i \in \{0, 1, \ldots, r\}$, there exists a formula (circuit) Φ_i such that $|\Phi_i| \leq 9r^2 \cdot |\Gamma|$, $\mathrm{depth}(\Phi_i) \leq \mathrm{depth}(\Gamma)$ and Φ_i computes the polynomial $\frac{\partial^i P}{(\partial y)^i}(\mathbf{x}, y)$.*

Given an irreducible polynomial $g(\mathbf{x}, y)$ and a polynomial $P(\mathbf{x}, y)$ that is divisible by g, it will be useful for us to find a polynomial $D(\mathbf{x}, y)$ that is divisible by g and it is also square-free with respect to g, that is, $g(\mathbf{x}, y) \nmid \frac{\partial D}{\partial y}(\mathbf{x}, y)$. The next lemma shows that we can find such a polynomial efficiently.

Lemma 2.6. *Let $g(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be an irreducible polynomial that divides a polynomial $P(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$, where $\deg_y(P) \leq r$ and let Γ be a formula computing $P(\mathbf{x}, y)$. Then, there exists a formula Δ that computes a polynomial $D(\mathbf{x}, y)$ such that $g(\mathbf{x}, y) \mid D(\mathbf{x}, y)$, $g(\mathbf{x}, y) \nmid \frac{\partial D}{\partial y}(\mathbf{x}, y)$, $|\Delta| \leq 9r^2 \cdot |\Gamma|$ and $\mathrm{depth}(\Delta) \leq \mathrm{depth}(\Gamma)$. Moreover, the output gate of Δ is an addition gate and for each variable $z \in \{\mathbf{x}, y\}$, we have that $\deg_z(D) \leq \deg_z(P)$.*

The following observation will be very useful to convert small depth formulas into formulas with fanning bounded by 2.

Observation 2.7. *Any formula Φ of size s and depth d, without restrictions on the in-degree of any of its gates, can be computed by a formula Ψ of size s and depth $d \cdot \log(s)$, where each gate has in-degree 2.*

To see that this observation is true, just replace each addition (multiplication) gate of in-degree t by a balanced formula of size t made only with addition (multiplication) gates. Since $t \leq s$, and a balanced formula of size t has depth $\log t$, we have that each gate will be replaced by a formula of depth at most $\log s$. The replacement by a balanced formula clearly does not change the computation neither the size of the formula, and the depth increases by a multiplicative factor of $\log s$, as we wanted.

2.3 Reversal of Polynomials

In this section, we define a very useful operation for polynomials, which serves as a crucial tool in the proof of our main theorem. This operation, which we call *reversal*, simply maps a polynomial $P(\mathbf{x}, y) \equiv \sum_{i=0}^{r} P_i(\mathbf{x}) y^i$, with $P_r(\mathbf{x}) \cdot P_0(\mathbf{x}) \not\equiv 0$, to $\tilde{P}(\mathbf{x}) \equiv \sum_{i=0}^{r} P_i(\mathbf{x}) y^{r-i}$.

The restriction that $P_r(\mathbf{x}) \cdot P_0(\mathbf{x}) \not\equiv 0$ is needed in this paper because it preserves irreducibility, as we will see in Lemma 2.9 and Corollary 2.10. We begin by showing that the reversal can be computed almost as efficiently as the original polynomial.

Lemma 2.8 (Reversal Lemma)**.** *Let* $P(\mathbf{x}, y) \equiv \sum_{i=0}^{r} y^i P_i(\mathbf{x})$ *be a polynomial computed by a formula (circuit)* Γ*, where* $P_r(\mathbf{x}) \cdot P_0(\mathbf{x}) \not\equiv 0$*. Let* $\tilde{P}(\mathbf{x}, y) \equiv \sum_{i=0}^{r} y^{r-i} P_i(\mathbf{x})$ *be its reversal. There exists a formula (circuit)* Δ *computing* $\tilde{P}$ *such that* $|\Delta| = 8r^2 \cdot |\Gamma|$*.*

We now connect the reversal operation to divisibility and irreducibility of polynomials.

Lemma 2.9 (Divisibility with Reversals)**.** *Let* $P(\mathbf{x}, y) \equiv \sum_{i=0}^{r} y^i P_i(\mathbf{x})$*, with* $P_r(\mathbf{x}) \cdot P_0(\mathbf{x}) \not\equiv 0$ *and* $f(\mathbf{x}, y) \equiv \sum_{i=0}^{k} y^i f_i(\mathbf{x})$*, with* $f_k(\mathbf{x}) \cdot f_0(\mathbf{x}) \not\equiv 0$*, be two polynomials. In addition, let* $\tilde{P}(\mathbf{x}, y) \equiv \sum_{i=0}^{r} y^{r-i} P_i(\mathbf{x})$ *and* $\tilde{f}(\mathbf{x}, y) \equiv \sum_{i=0}^{k} y^{k-i} f_i(\mathbf{x})$ *be their reversals. Then, we have that*

$$f \mid P \iff \tilde{f} \mid \tilde{P}.$$

Since divisibility is preserved by taking reversals, we have the following corollary:

Corollary 2.10 (Irreducibility of Reversals)**.** *Let* $P(\mathbf{x}, y) \equiv \sum_{i=0}^{r} y^i P_i(\mathbf{x})$*, with* $P_r(\mathbf{x}) \cdot P_0(\mathbf{x}) \not\equiv 0$*, be an irreducible polynomial in* $\mathbb{F}[\mathbf{x}, y]$*. In addition, let* $\tilde{P}(\mathbf{x}, y) \equiv \sum_{i=0}^{r} y^{r-i} P_i(\mathbf{x})$ *be its reversal. Then, we have that*

$$P \text{ is irreducible} \iff \tilde{P} \text{ is irreducible}.$$

Proof. From Lemma 2.9, we get that P is reducible iff $\tilde{P}$ is reducible. This implies the statement of the corollary. $\square$

Another useful property of reversals is that if two univariate polynomials do not share a common root, then their reversals will not share any root either. The proof of this fact is in the following lemma:

Lemma 2.11. *If $f(x), g(x) \in \mathbb{F}[x]$ do not share any common roots, then their reversals $\tilde{f}(x), \tilde{g}(x)$ do not share any roots either.*

3 Properly Splitting Variable Restrictions

In this section, we show how to find a shift of the input polynomial so that the restriction of $\mathbf{x}$ to $\mathbf{0}$ of the shifted polynomial has no repeated roots. In addition, we will be interested in the greatest common divisor of two polynomials with respect to one of their variables. That is, if two polynomials share a common factor involving a specific variable. For this purpose, we define the resultant of two polynomials $f(\mathbf{x}, y)$ and $g(\mathbf{x}, y)$ with respect to variable y and discuss some of its properties.

In addition, if two polynomials $f(\mathbf{x}, y), g(\mathbf{x}, y)$ share no common factor involving variable y we prove a lemma on restrictions of the $\mathbf{x}$ variables that preserve this property. We begin with the definition of *properly splitting* of the restriction of some variables.

Definition 3.1 (Properly Splitting Restrictions)**.** *Let $\mathbf{x} = (x_1, \ldots, x_n)$, where $n \geq 1$, and let $f(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be an irreducible polynomial such that $\deg_y(f) \geq 1$. In addition, let $g(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be a polynomial with $\deg_y(g) \geq 1$ that is not divisible by $f(\mathbf{x}, y)$. We say that $\mathbf{c} \in \mathbb{F}^n$* properly splits *$f(\mathbf{x}, y)$ with respect to $g(\mathbf{x}, y)$ if the following conditions hold:*

(i) $f(\mathbf{c}, y)$ is a polynomial with exactly $\deg_y(f)$ distinct *roots in $\mathbb{F}$ and*

(ii) $f(\mathbf{c}, y)$ and $g(\mathbf{c}, y)$ share no common roots.

We now go on to define and state some useful properties of the resultant of two polynomials. All of the definitions and propositions below are taken from the book [CLO06, Chapter 3]. We begin by defining the Sylvester matrix of two polynomials.

Definition 3.2 (Sylvester Matrix, from [CLO06], §3.6)**.** *Let $f(\mathbf{x}, y) = \sum_{i=0}^{\ell} f_i(\mathbf{x}) y^i$ and $g(\mathbf{x}, y) = \sum_{i=0}^{m} g_i(\mathbf{x}) y^i$ be two polynomials in $\mathbb{F}[\mathbf{x}, y]$ such that $\ell, m > 0$ and $f_\ell(\mathbf{x}) \cdot g_m(\mathbf{x}) \not\equiv 0$. Then, we* *define the* Sylvester Matrix *of f, g with respect to y as the following $(m+\ell)\times(m+\ell)$ matrix:*

$$\mathrm{Syl}(f,g,y) = \begin{pmatrix} f_\ell & & & & g_m & & & \\ f_{\ell-1} & f_\ell & & & g_{m-1} & g_m & & \\ f_{\ell-2} & f_{\ell-1} & \ddots & & g_{m-2} & g_{m-1} & \ddots & \\ \vdots & \vdots & \ddots & f_\ell & \vdots & \vdots & \ddots & g_m \\ f_0 & f_1 & & f_{\ell-1} & g_0 & g_1 & & g_{m-1} \\ & f_0 & & \vdots & & g_0 & & \vdots \\ & & \ddots & & & & \ddots & \\ & & & f_0 & & & & g_0 \end{pmatrix}$$

where the first m columns have as entries the coefficients of f and the last ℓ columns have as entries the coefficients of g (when seen as polynomials in $\mathbb{F}[\mathbf{x}][y]$). All other entries are zero.

We are now ready to define the resultant of two polynomials.

Definition 3.3 (Resultants, from §3.6 of [CLO06])**.** *Given polynomials $f, g \in \mathbb{F}[\mathbf{x}, y]$ of positive degree in y, we define the* Resultant *of f and g with respect to y, written $\mathrm{Res}(f,g,y)$, as the determinant of the Sylvester matrix. That is,*

$$\mathrm{Res}(f,g,y) \equiv \det(\mathrm{Syl}(f,g,y)).$$

Observation 3.4. *From the definition above, we have that $\mathrm{Res}(f,g,y)$ is a polynomial in $\mathbb{F}[\mathbf{x}]$ with degree bounded by $2\deg(f)\deg(g)$, since each term in the expression of the determinant will be a product of m of the coefficients f_i by ℓ of the coefficients g_j.*

A very useful property of the resultant is the following:

Proposition 3.5 (Main Property of Resultants, Proposition 1 in §3.6 of [CLO06])**.** *Let $f, g \in \mathbb{F}[\mathbf{x}, y]$ be polynomials with positive degree in y. Then, $\mathrm{Res}(f,g,y) \equiv 0$ if, and only if, f and g have a common factor in $\mathbb{F}[\mathbf{x}, y]$ which has positive degree in y.*

With the definitions and facts above, we are now ready to state and prove the main lemma of this section. This lemma tells us that the set of restrictions that properly split an irreducible polynomial $f(\mathbf{x}, y)$ with respect to a polynomial $g(\mathbf{x}, y)$ that is not divisible by $f(\mathbf{x}, y)$ is the complement of an algebraic set. This implies that a random restriction of the variables $\mathbf{x}$ will properly split $f(\mathbf{x}, y)$ with respect to $g(\mathbf{x}, y)$.

Lemma 3.6. *Let $\mathbf{x} = (x_1, \dots, x_n)$, where $n \geq 1$ and $f(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be an irreducible polynomial such that $\deg_y(f) \geq 1$. In addition, let $g(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be a polynomial with $\deg_y(g) \geq 1$ that is not divisible by $f(\mathbf{x}, y)$. Then, there exists a nonzero polynomial $G(\mathbf{x})$ with $\deg(G) \leq 2\deg(f)^2 + 2\deg(f)\deg(g)$ for which the following holds: for any value $\mathbf{c} \in \mathbb{F}^n$ such that $G(\mathbf{c}) \neq 0$, we have that $\mathbf{c}$ properly splits $f(\mathbf{x}, y)$ with respect to $g(\mathbf{x}, y)$.*

Proof. Let $f(\mathbf{x}, y) = \sum_{i=0}^{k} f_i(\mathbf{x}) y^i$, with $f_k(\mathbf{x}) \not\equiv 0$ and $f'(\mathbf{x}, y) \equiv \dfrac{\partial f}{\partial y}(\mathbf{x}, y)$. If $\deg_y(f) > 1$, we have that f' also has positive degree in y. Now, by irreducibility of f and by proposition 3.5, we have that $p(\mathbf{x}) \equiv \mathrm{Res}(f, f', y)$ is a nonzero polynomial. Since $f \nmid g$ and f is irreducible, f and g share no common factors, and hence we also must have that $q(\mathbf{x}) \equiv \mathrm{Res}(f, g, y) \not\equiv 0$. Therefore, we have that

$$f_k(\mathbf{x})p(\mathbf{x})q(\mathbf{x}) \not\equiv 0,$$

as this is the product of nonzero polynomials. Notice that

$$\deg(f_k(\mathbf{x})p(\mathbf{x})q(\mathbf{x})) = \deg(f_k(\mathbf{x})) + \deg(p(\mathbf{x})) + \deg(q(\mathbf{x})) \leq 2\deg(f)^2 + 2\deg(f)\deg(g)$$

Notice that for any $\mathbf{c} \in \mathbb{F}^n$ such that $f_k(\mathbf{c})p(\mathbf{c})q(\mathbf{c}) \neq 0$, this implies that $f(\mathbf{c}, y)$ still has degree k, that $\mathrm{Res}(f(\mathbf{c}, y), g(\mathbf{c}, y), y) \neq 0$, and that $\mathrm{Res}(f(\mathbf{c}, y), f'(\mathbf{c}, y), y) \neq 0$. Thus, by proposition 3.5 we have that $f(\mathbf{c}, y)$ and $f'(\mathbf{c}, y)$ have no common factors in $\mathbb{F}[y]$, and therefore no common roots. Thereby, $f(\mathbf{c}, y)$ has no repeated root, which implies that $f(\mathbf{c}, y)$ has exactly $k = \deg_y(f)$ distinct roots. Similarly $f(\mathbf{c}, y)$ and $g(\mathbf{c}, y)$ have no common roots.

If $\deg_y(f) = 1$, then $\deg_y(f) = 1$ implies that any restriction of the form $f(\mathbf{c}, y)$ such that $f_1(\mathbf{c}) \neq 0$ will yield exactly one root. And this root must be distinct from the roots of $g(\mathbf{c}, y)$, for $\mathrm{Res}(f(\mathbf{c}, y), g(\mathbf{c}, y), y) \neq 0$.

Letting $G(\mathbf{x}) \equiv f_k(\mathbf{x})p(\mathbf{x})q(\mathbf{x})$ completes the proof. □

4 Standard Forms and Approximation Polynomials

In this section we define the notion of standard forms in $\mathbb{F}(\mathbf{x})[y]$, that is, the ring of polynomials on the variable y with coefficients being rational functions on the variables $\mathbf{x}$. We also define the approximation polynomial of a standard form. These concepts will be useful when factoring a polynomial $P(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$, since our factorization procedure will use standard forms to obtain the factors of $P(\mathbf{x}, y)$ that depend of the variable y. We begin with the following definition:

Definition 4.1 (Standard Form and Approximation Polynomials). *We say that $f(\mathbf{x}, y) \in \mathbb{F}(\mathbf{x})[y]$ is in* standard form *if*

$$f(\mathbf{x}, y) \equiv \frac{1}{1 - g(\mathbf{x})} \cdot \sum_{i=0}^{k} f_i(\mathbf{x}) y^i,$$

where $f_i(\mathbf{x}), g(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$, $f_k(\mathbf{x}) \not\equiv 0$ and $g(\mathbf{0}) = 0$. Moreover, we will say that f is in monic standard form *if $f_k(\mathbf{x}) \equiv 1 - g(\mathbf{x})$. For a given parameter $m \in \mathbb{N}$, we define the* approximation polynomial *of the standard form f to degree m, as the polynomial $\psi_{f,m}(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ given by*

$$\psi_{f,m}(\mathbf{x}, \mathbf{y}) = (1 + g(\mathbf{x}) + \ldots + g(\mathbf{x})^m) \cdot \sum_{i=0}^{k} f_i(\mathbf{x}) y^i.$$

In order to state some useful properties of approximation polynomials, we will need to extend the definition of reversals to standard forms.

Definition 4.2. *Let $f(\mathbf{x}, y)$ be a standard form as above, with the additional condition that $f_0(\mathbf{x}) \not\equiv 0$. We define the reversal of $f(\mathbf{x}, y)$ as the following standard form:*

$$\tilde{f}(\mathbf{x}, y) \equiv \frac{1}{1 - g(\mathbf{x})} \cdot \sum_{i=0}^{k} f_i(\mathbf{x}) y^{k-i}.$$

The following observations about standard forms reveal much of its usefulness when factoring a polynomial.

Observation 4.3. *If $f(\mathbf{x}, y) \in \mathbb{F}(\mathbf{x})[y]$ is in standard form as above, notice that the following holds for all $m \in \mathbb{N}$:*

1. $H^{\mathbf{x}}_{\leq m}[(1 - g(\mathbf{x})) \cdot \psi_{f,m}(\mathbf{x}, y)] \equiv H^{\mathbf{x}}_{\leq m}[(1 - g(\mathbf{x})) \cdot f(\mathbf{x}, y)]$.
2. *If* $m \geq \deg((1 - g(\mathbf{x})) \cdot f(\mathbf{x}, y))$, *we have:*

$$H^{\mathbf{x}}_{\leq m}[(1 - g(\mathbf{x})) \cdot \psi_{f,m}(\mathbf{x}, y)] \equiv (1 - g(\mathbf{x})) \cdot f(\mathbf{x}, y).$$

3. $H^{\mathbf{x}}_{\leq m}[\psi_{\tilde{f},m}(\mathbf{x}, y)] \equiv H^{\mathbf{x}}_{\leq m}\left[\widetilde{\psi}_{f,m}(\mathbf{x}, y)\right]$.
4. *If* $h(\mathbf{x}, y) \equiv f(\mathbf{x}, y + \gamma)$, *where* $\gamma \in \mathbb{F}$, *we have that* $h(\mathbf{x}, y)$ *is also a standard form and*

$$H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x}, y + \gamma)] \equiv H^{\mathbf{x}}_{\leq m}[\psi_{h,m}(\mathbf{x}, y)].$$

Proof. 1. Notice that

$$\begin{aligned} H^{\mathbf{x}}_{\leq m}\left[(1 - g(\mathbf{x})) \cdot \psi_{f,m}(\mathbf{x}, y)\right] &\equiv H^{\mathbf{x}}_{\leq m}\left[(1 - g(\mathbf{x})) \cdot (1 + g(\mathbf{x}) + \ldots + g(\mathbf{x})^m) \cdot \sum_{i=0}^{k} f_i(\mathbf{x}) y^i\right] \\ &\equiv H^{\mathbf{x}}_{\leq m}\left[(1 - g(\mathbf{x})^{m+1}) \cdot \sum_{i=0}^{k} f_i(\mathbf{x}) y^i\right] \\ &\equiv H^{\mathbf{x}}_{\leq m}\left[\sum_{i=0}^{k} f_i(\mathbf{x}) y^i\right] \equiv H^{\mathbf{x}}_{\leq m}\left[(1 - g(\mathbf{x})) \cdot f(\mathbf{x}, y)\right] \end{aligned}$$

2. Given item 1 and the fact that $H^{\mathbf{x}}_{\leq m}[p(\mathbf{x}, y)] \equiv p(\mathbf{x}, y)$ for all $m \geq \deg(p)$, this part follows from the fact that

$$H^{\mathbf{x}}_{\leq m}\left[(1 - g(\mathbf{x})) \cdot f(\mathbf{x}, y)\right] \equiv (1 - g(\mathbf{x})) \cdot f(\mathbf{x}, y).$$

3. This item follows from the fact that

$$\psi_{\tilde{f},m}(\mathbf{x}, y) \equiv (1 + g(\mathbf{x}) + \ldots + g(\mathbf{x})^m) \cdot \sum_{i=0}^{k} f_i(\mathbf{x}) y^{k-i} \equiv \widetilde{\psi}_{f,m}(\mathbf{x}, y)$$

4. Notice that

$$\begin{aligned}
h(\mathbf{x}, y) \equiv f(\mathbf{x}, y+\gamma) &\equiv \frac{1}{1-g(\mathbf{x})} \cdot \sum_{i=0}^{k} f_i(\mathbf{x})(y+\gamma)^i \\
&\equiv \frac{1}{1-g(\mathbf{x})} \cdot \sum_{i=0}^{k} \left(\sum_{j=i}^{k} \binom{j}{i} \gamma^{j-i} f_j(\mathbf{x}) \right) y^i \\
&\equiv \frac{1}{1-g(\mathbf{x})} \cdot \sum_{i=0}^{k} h_i(\mathbf{x}) y^i, \quad \text{where} \quad h_i(\mathbf{x}) \equiv \sum_{j=i}^{k} \binom{j}{i} \gamma^{j-i} f_j(\mathbf{x}).
\end{aligned}$$

Since each term $h_i(\mathbf{x})$ is in $\mathbb{F}[\mathbf{x}]$, this proves that $h(\mathbf{x}, y)$ is also a standard form. Moreover, from the equalities above, we have that

$$\begin{aligned}
\psi_{f,m}(\mathbf{x}, y+\gamma) &\equiv (1 + g(\mathbf{x}) + \ldots + g(\mathbf{x})^m) \cdot \sum_{i=0}^{k} f_i(\mathbf{x})(y+\gamma)^i \\
&\equiv (1 + g(\mathbf{x}) + \ldots + g(\mathbf{x})^m) \cdot \sum_{i=0}^{k} \left(\sum_{j=i}^{k} \binom{j}{i} \gamma^{j-i} f_j(\mathbf{x}) \right) y^i \\
&\equiv (1 + g(\mathbf{x}) + \ldots + g(\mathbf{x})^m) \cdot \sum_{i=0}^{k} h_i(\mathbf{x}) y^i \equiv \psi_{h,m}(\mathbf{x}, y).
\end{aligned}$$

From this equality the relation in item 4 follows.

□

5 Approximating the Roots of a Polynomial

In this section, we proceed in a similar way as in [DSY09] and find approximations of the roots of a polynomial $P(\mathbf{x}, y)$ up to degree t. That is, as we defined in the introduction, we find polynomials $q_t(\mathbf{x})$ such that $H^{\mathbf{x}}_{\leq t}[P(\mathbf{x}, q_t(\mathbf{x}))] \equiv 0$. Moreover, we observe that under certain conditions on the polynomial $P(\mathbf{x}, y)$ these roots are well-defined and unique given their constant coefficient. This uniqueness condition will be useful because it will allow us to construct any factor of $P(\mathbf{x}, y)$ through the lifting procedure, since a factor $f(\mathbf{x}, y)$ of $P(\mathbf{x}, y)$ will share some of the roots of $P(\mathbf{x}, y)$. We begin with the approximation lemma:

Lemma 5.1 (Approximation Lemma). *Let $P(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$, $P'(\mathbf{x}, y) \equiv \frac{\partial P}{\partial y}(\mathbf{x}, y)$ and $\mu \in \mathbb{F}$ be such that $P(\mathbf{0}, \mu) = 0$ but $P'(\mathbf{0}, \mu) = \xi \neq 0$. Then, for each $t \geq 0$, there exists a* **unique** *polynomial $q_t(\mathbf{x})$ s.t.* $\deg(q_t) \leq t$, $q_t(\mathbf{0}) = \mu$ *and*

$$H^{\mathbf{x}}_{\leq t}[P(\mathbf{x}, q_t(\mathbf{x}))] \equiv 0.$$

Moreover, if P can be computed by a formula (circuit) Γ such that its output gate is an addition gate, there is a formula (circuit) Φ_t for the polynomial $q_t(\mathbf{x})$ such that the output gate of Φ_t is *an addition gate,* $\mathrm{depth}(\Phi_t) \le \mathrm{depth}(\Gamma) + 2$ *and*

$$|\Phi_t| \le 200(tr)^2 \binom{t+r+1}{r+1} \cdot |\Gamma|.$$

If we require the in-degree of the formula (circuit) to be 2, then the size of Φ_t does not change, and $\mathrm{depth}(\Phi_t) \le \mathrm{depth}(\Gamma) + 5r\log(t)$*.*

Proof. The proof of uniqueness of $q_t(\mathbf{x})$ and the construction of a small formula computing $q_t(\mathbf{x})$ are done by induction on t. For the rest of the proof, let $P(\mathbf{x}, y) = \sum_{i=0}^{r} C_i(\mathbf{x}) y^i$, $\tilde{C}_i(\mathbf{x}) = C_i(\mathbf{x}) - C_i(\mathbf{0})$ and $\mathbf{z} = (z_0, \ldots, z_r)$.

We will conclude the inductive proof with two steps:

Step 1: existence and uniqueness of $q_t(\mathbf{x})$.

For $t = 0$, note that $q_0(\mathbf{x}) \equiv \mu$ is the only solution. Hence, step 1 is true for $t = 0$. Assume now that the existence and uniqueness of $q_t(\mathbf{x})$ are true for $t - 1$, where $t \ge 1$.

Firstly, notice that any polynomial $q_t(\mathbf{x})$ that satisfies $0 \equiv H^{\mathbf{x}}_{\le t}[P(\mathbf{x}, q_t(\mathbf{x}))]$ must satisfy $0 \equiv H^{\mathbf{x}}_{\le t-1}[P(\mathbf{x}, q_t(\mathbf{x}))]$ and therefore we must have that $q_t(\mathbf{x}) \equiv g(\mathbf{x}) + q_{t-1}(\mathbf{x})$, where $g(\mathbf{x})$ is homogeneous of degree t (by the uniqueness of $q_{t-1}(\mathbf{x})$).

Hence, we have

$$\begin{aligned}
&0 \equiv H^{\mathbf{x}}_{\le t}[P(\mathbf{x}, q_t(\mathbf{x}))] \equiv H^{\mathbf{x}}_{\le t}[P(\mathbf{x}, q_{t-1}(\mathbf{x}) + g(\mathbf{x}))] \equiv H^{\mathbf{x}}_{\le t}\left[\sum_{i=0}^{r} C_i(\mathbf{x})(q_{t-1}(\mathbf{x}) + g(\mathbf{x}))^i\right] \\
&\equiv H^{\mathbf{x}}_{\le t}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + H^{\mathbf{x}}_{\le t}\left[\sum_{i=0}^{r} i C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^{i-1} g(\mathbf{x})\right] \\
&\equiv H^{\mathbf{x}}_{t}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + \sum_{i=0}^{r} i C_i(\mathbf{0}) q_{t-1}(\mathbf{0})^{i-1} g(\mathbf{x}) \equiv H^{\mathbf{x}}_{t}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + \xi \cdot g(\mathbf{x}) \\
\iff &g(\mathbf{x}) \equiv -\frac{1}{\xi} \cdot H^{\mathbf{x}}_{t}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right]
\end{aligned}$$

By uniqueness of $q_{t-1}(\mathbf{x})$, we obtain that $g(\mathbf{x})$ is unique, and therefore, $q_t(\mathbf{x})$ must also be unique. Existence of $q_t(\mathbf{x})$ follows from the above and from existence of $q_{t-1}(\mathbf{x})$. This completes the induction of step 1.

Step 2: construction of a formula Φ_t for $q_t(\mathbf{x})$ based on the formula Γ computing $P(\mathbf{x}, y)$, such that $|\Phi_t| \le 200(tr)^2 \binom{t+r+1}{r+1} \cdot |\Gamma|$ and $\mathrm{depth}(\Phi_t) \le \mathrm{depth}(\Gamma) + 2$.

To construct Φ_t, we will first inductively construct a polynomial $Q_t(\mathbf{z}) \in \mathbb{F}[\mathbf{z}]$ such that $\deg_{\mathbf{z}}(Q_t) \le t$ and

$$H^{\mathbf{x}}_{\le t}[Q_t(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))] \equiv q_t(\mathbf{x}).$$

For $t = 0$, we can set $Q_0(\mathbf{z}) = \mu$. This way, we have that $\deg_{\mathbf{z}}(Q_0) = 0$ and

$$H_{\leq 0}[Q_0(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x}))] \equiv Q_0(\mathbf{z}) = \mu = q_0(\mathbf{x}).$$

Now, assume that the claim is true for $t-1$, where $t \geq 1$. Therefore, we have the polynomial $Q_{t-1}(\mathbf{z})$ with the given properties above. With this polynomial, we claim that

$$Q_t(\mathbf{z}) \equiv H^{\mathbf{x}}_{\leq t}\left[Q_{t-1}(\mathbf{z}) - \frac{1}{\xi} \cdot \sum_{i=0}^{r} (z_i + C_i(\mathbf{0})) \cdot Q_{t-1}(\mathbf{z})^i\right] \tag{1}$$

is a polynomial such that

$$q_t(\mathbf{x}) \equiv H^{\mathbf{x}}_{\leq t}[Q_t(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x}))].$$

Notice that

$$\begin{aligned}
&H^{\mathbf{x}}_{\leq t-1}[Q_t(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x}))] \equiv \\
&\equiv H_{\leq t-1}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x})) - \frac{1}{\xi} \cdot \sum_{i=0}^{r} C_i(\mathbf{x}) Q_{t-1}(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x}))^i\right] \\
&\equiv H^{\mathbf{x}}_{\leq t-1}[Q_{t-1}(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x}))] - \frac{1}{\xi} \cdot H^{\mathbf{x}}_{\leq t-1}[P(\mathbf{x}, Q_{t-1}(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x})))] \\
&\equiv q_{t-1}(\mathbf{x}) + 0 \equiv q_{t-1}(\mathbf{x})
\end{aligned}$$

Therefore, $H^{\mathbf{x}}_{\leq t}\left[Q_t(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x}))\right] \equiv g(\mathbf{x}) + q_{t-1}(\mathbf{x})$, where $g(\mathbf{x})$ is homogeneous of degree t (if it is nonzero). Based on (1), we have that

$$g(\mathbf{x}) \equiv H^{\mathbf{x}}_{t}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x})) - \frac{1}{\xi} \cdot \sum_{i=0}^{r} C_i(\mathbf{x}) \cdot Q_{t-1}(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x}))^i\right] \tag{2}$$

Now, notice that

$$\begin{aligned}
H^{\mathbf{x}}_{\leq t}\left[P(\mathbf{x}, Q_t(\tilde{C}_0(\mathbf{x}), \dots, \tilde{C}_r(\mathbf{x})))\right] &\equiv H^{\mathbf{x}}_{\leq t}\left[P(\mathbf{x}, q_{t-1}(\mathbf{x}) + g(\mathbf{x}))\right] \\
&\equiv H^{\mathbf{x}}_{\leq t}\left[\sum_{i=0}^{r} C_i(\mathbf{x})(q_{t-1}(\mathbf{x}) + g(\mathbf{x}))^i\right] \\
&\equiv H^{\mathbf{x}}_{\leq t}\left[\sum_{i=0}^{r} C_i(\mathbf{x})\left(q_{t-1}(\mathbf{x})^i + i q_{t-1}(\mathbf{x})^{i-1} g(\mathbf{x})\right)\right]
\end{aligned}$$

where the last equation is true because the terms with $g(\mathbf{x})^k$, where $k > 1$, all have degree larger than t. Since we are only looking at parts of degree $\leq t$, we can refine the equation above even further to obtain

$$
\begin{aligned}
&H^{\mathbf{x}}_{\leq t}\left[\sum_{i=0}^{r} C_i(\mathbf{x})\left(q_{t-1}(\mathbf{x})^i + i q_{t-1}(\mathbf{x})^{i-1} g(\mathbf{x})\right)\right] \\
&\equiv H^{\mathbf{x}}_{\leq t}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + H^{\mathbf{x}}_{\leq t}\left[\sum_{i=0}^{r} i C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^{i-1} g(\mathbf{x})\right] \\
&\equiv H_{\leq t-1}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + H_t\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + H_t\left[\sum_{i=0}^{r} i C_i(\mathbf{0}) q_{t-1}(\mathbf{0})^{i-1} g(\mathbf{x})\right] \\
&\equiv 0 + H_t\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + g(\mathbf{x}) \cdot \sum_{i=0}^{r} i C_i(\mathbf{0}) \mu^{i-1} \\
&\equiv H_t\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + g(\mathbf{x}) \cdot \xi \qquad (3)
\end{aligned}
$$

where $H_{\leq t-1}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] \equiv H^{\mathbf{x}}_{\leq t-1}[P(\mathbf{x}, q_{t-1}(\mathbf{x}))] \equiv 0$ by our induction hypothesis and we know that $\sum_{i=0}^{r} i C_i(\mathbf{0}) \mu^{i-1} = P'(\mathbf{0}, \mu) = \xi$.

By our induction hypothesis,

$$
\begin{aligned}
q_{t-1}(\mathbf{x}) &\equiv H_{\leq t-1}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))\right] \\
&\equiv H^{\mathbf{x}}_{\leq t}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))\right] - H^{\mathbf{x}}_{t}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))\right].
\end{aligned}
$$

Setting $A(\mathbf{x}) \equiv H^{\mathbf{x}}_{\leq t}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))\right]$ and $B(\mathbf{x}) \equiv H^{\mathbf{x}}_{t}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))\right]$, and substituting this expression for $q_{t-1}(\mathbf{x})$ in (3), we get:

$$
\begin{aligned}
&H_t\left[\sum_{i=0}^{r} C_i(\mathbf{x}) q_{t-1}(\mathbf{x})^i\right] + g(\mathbf{x})\cdot\xi \\
&\equiv H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} C_i(\mathbf{x})\left(A(\mathbf{x}) - B(\mathbf{x})\right)^i\right] + g(\mathbf{x})\cdot\xi \\
&\equiv H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) A(\mathbf{x})^i\right] - H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} iC_i(\mathbf{x}) A(\mathbf{x})^{i-1} B(\mathbf{x})\right] + g(\mathbf{x})\cdot\xi \\
&\equiv H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) A(\mathbf{x})^i\right] - H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} iC_i(\mathbf{0}) A(\mathbf{0})^{i-1} B(\mathbf{x})\right] + g(\mathbf{x})\cdot\xi \\
&\equiv H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) A(\mathbf{x})^i\right] - \left(\sum_{i=0}^{r} iC_i(\mathbf{0}) \mu^{i-1}\right)\cdot B(\mathbf{x}) + g(\mathbf{x})\cdot\xi \\
&\equiv H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) A(\mathbf{x})^i\right] - \xi\cdot B(\mathbf{x}) + g(\mathbf{x})\cdot\xi
\end{aligned} \tag{4}
$$

Where in the equations above we used the facts that $B(\mathbf{x})$ is a homogeneous polynomial of degree t, that $A(\mathbf{0}) = H^{\mathbf{x}}_{\leq t}\left[Q_{t-1}(\tilde{C}_0(\mathbf{0}),\ldots,\tilde{C}_r(\mathbf{0}))\right] = q_{t-1}(\mathbf{0}) = \mu$ and that $\sum_{i=0}^{r} iC_i(\mathbf{0})\mu^{i-1} = P'(\mathbf{0},\mu) = \xi$.

Therefore, substituting the expressions for $A(\mathbf{x}), B(\mathbf{x})$, together with the expression for $g(\mathbf{x})$ from equation (2), we have:

$$
\begin{aligned}
&H^{\mathbf{x}}_{\leq t}\left[P(\mathbf{x}, Q_t(\tilde{C}_0(\mathbf{x}),\ldots,\tilde{C}_r(\mathbf{x})))\right] \\
&\equiv H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) A(\mathbf{x})^i\right] - \xi\cdot B(\mathbf{x}) + g(\mathbf{x})\cdot\xi \\
&\equiv H_t^{\mathbf{x}}\left[\sum_{i=0}^{r} C_i(\mathbf{x}) H^{\mathbf{x}}_{\leq t}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}),\ldots,\tilde{C}_r(\mathbf{x}))\right]^i\right] - \xi\cdot H_t^{\mathbf{x}}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}),\ldots,\tilde{C}_r(\mathbf{x}))\right] + \\
&\xi\cdot H_t^{\mathbf{x}}\left[Q_{t-1}(\tilde{C}_0(\mathbf{x}),\ldots,\tilde{C}_r(\mathbf{x})) - \frac{1}{\xi}\cdot\sum_{i=0}^{r} C_i(\mathbf{x})\cdot Q_{t-1}(\tilde{C}_0(\mathbf{x}),\ldots,\tilde{C}_r(\mathbf{x}))^i\right] \equiv 0
\end{aligned}
$$

As we wanted. This finishes the inductive step and shows that $q_t(\mathbf{x}) \equiv H^{\mathbf{x}}_{\leq t}\left[Q_t(\tilde{C}_0(\mathbf{x}),\ldots,\tilde{C}_r(\mathbf{x}))\right]$.

Now, we only need to construct a formula Φ_t computing $q_t(\mathbf{x})$. Since $Q_t(\mathbf{z})$ is a polynomial of degree $\leq t$ in $r+1$ variables, its sparse (depth 2) representation has $\leq \binom{t+r+1}{r+1}$ monomials. Hence, $Q_t(\mathbf{z})$ can be computed by a formula Δ_t of depth 2 and size $\leq 2tr^2\binom{t+r+1}{r+1}$. If we require the formula to have in-degree 2, then Observation 2.7 implies that Δ_t can be computed by a formula of the same size and depth $\leq \log(2tr^2\binom{t+r+1}{r+1}) \leq 3r\log(t)$.

Since $P(\mathbf{x}, y)$ can be computed by a formula of size $|\Gamma|$, by Lemma 2.4 we have that each $C_i(\mathbf{x})$ can be computed by a formula Γ_i such that $|\Gamma_i| \leq 3r \cdot |\Gamma|$. Since $\tilde{C}_i(\mathbf{x}) \equiv C_i(\mathbf{x}) - C_i(\mathbf{0})$, there exists a formula $\tilde{\Gamma}_i$ computing $\tilde{C}_i(\mathbf{x})$ of size $|\tilde{\Gamma}_i| \leq 5r \cdot |\Gamma|$. Hence, by composing many copies of the formulas $\tilde{\Gamma}_i$ with the formula Δ_t, and by noticing that the polynomial Q_t has degree $\leq t$, we obtain a formula Ψ_t of size $\leq 2t|\Delta_t| \cdot (5r^2 \cdot |\Gamma|)$ computing the composition polynomial $Q_t(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))$. Because Ψ_t is a composition, the depth of Ψ_t is given by

$$\operatorname{depth}(\Psi_t) = \operatorname{depth}(\Delta_t) + \max_{0 \leq i \leq r}\left(\operatorname{depth}(\tilde{\Gamma}_i)\right) = \operatorname{depth}(\Delta_t) + \operatorname{depth}(\Gamma) = \operatorname{depth}(\Gamma) + 2.$$

Moreover, the topmost gate of Ψ_t is an addition gate.

In the case of formulas with in-degree bounded by 2, the depth of the formulas (circuits) $\tilde{\Gamma}_i$ are bounded by $\operatorname{depth}(\Gamma) + 2\log r$, in which case $\operatorname{depth}(\Psi_t) = \operatorname{depth}(\Delta_t) + \operatorname{depth}(\Gamma) + 2\log r \leq \operatorname{depth}(\Gamma) + 4r\log t$.

Since $q_t(\mathbf{x}) \equiv H^{\mathbf{x}}_{\leq t}\left[Q_t(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))\right]$ and $Q_t(\tilde{C}_0(\mathbf{x}), \ldots, \tilde{C}_r(\mathbf{x}))$ can be computed by formula Ψ_t of size $|\Psi_t| \leq 2t|\Delta_t| \cdot (5r^2 \cdot |\Gamma|)$ and with a topmost addition gate, Lemma 2.3 implies that there is a formula Φ_t computing $q_t(\mathbf{x})$ such that $\operatorname{depth}(\Phi_t) = \operatorname{depth}(\Psi_t) = \operatorname{depth}(\Gamma) + 2$ and

$$|\Phi_t| \leq 10t \cdot |\Psi_t| \leq 100tr^2|\Delta_t| \cdot |\Gamma| \leq 200(tr)^2\binom{t+r+1}{r+1} \cdot |\Gamma|$$

as we wanted.

In the case of formulas with in-degree bounded by 2, we would have $\operatorname{depth}(\Phi_t) = \operatorname{depth}(\Psi_t) + 2\log t = \operatorname{depth}(\Gamma) + 5r\log t$. $\square$

Now that we proved that any root of a polynomial $P(\mathbf{x}, y)$ of small individual degree computed by a small formula can be approximated by a small formula, the next corollary uses the uniqueness of the approximation of the root to show that the same is true for any factor of $P(\mathbf{x}, y)$.

Corollary 5.2. *Let $P(\mathbf{x}, y)$ and $\mu \in \mathbb{F}$ be defined as in Lemma 5.1 and for each $t \in \mathbb{N}_0$, let $q_t(\mathbf{x})$ be the unique polynomial obtained from Lemma 5.1. If $h(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ is such that $h(\mathbf{0}, \mu) = 0$, $\frac{\partial h}{\partial y}(\mathbf{0}, \mu) \neq 0$ and there exist $t \in \mathbb{N}$ and $Q(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ such that*

$$H^{\mathbf{x}}_{\leq t}[P(\mathbf{x}, y)] \equiv H^{\mathbf{x}}_{\leq t}[h(\mathbf{x}, y) \cdot Q(\mathbf{x}, y)], \tag{5}$$

then the polynomial $q_t(\mathbf{x})$ also satisfies

$$H^{\mathbf{x}}_{\leq t}[h(\mathbf{x}, q_t(\mathbf{x}))] \equiv 0, \ \forall t \geq 0.$$

Proof. Since μ is a root of $h(\mathbf{0}, y)$ and $\frac{\partial h}{\partial y}(\mathbf{0}, \mu) \neq 0$, Lemma 5.1 implies that there exists a unique $g_t(\mathbf{x})$ such that $H^{\mathbf{x}}_{\leq t}[h(\mathbf{x}, g_t(\mathbf{x}))] \equiv 0$. From equation (5), we must also have that

$H^{\mathbf{x}}_{\le t}[P(\mathbf{x}, g_t(\mathbf{x}))] \equiv 0$, since

$$\begin{aligned} H^{\mathbf{x}}_{\le t}[P(\mathbf{x}, y)] &\equiv H^{\mathbf{x}}_{\le t}[h(\mathbf{x}, y) \cdot Q(\mathbf{x}, y)] \Rightarrow \\ H^{\mathbf{x}}_{\le t}[P(\mathbf{x}, g_t(\mathbf{x}))] &\equiv H^{\mathbf{x}}_{\le t}[h(\mathbf{x}, g_t(\mathbf{x})) \cdot Q(\mathbf{x}, g_t(\mathbf{x}))] \\ &\equiv H^{\mathbf{x}}_{\le t}[H^{\mathbf{x}}_{\le t}[h(\mathbf{x}, g_t(\mathbf{x}))] \cdot Q(\mathbf{x}, g_t(\mathbf{x}))] \\ &\equiv H^{\mathbf{x}}_{\le t}[0 \cdot Q(\mathbf{x}, g_t(\mathbf{x}))] \equiv 0. \end{aligned}$$

The uniqueness of Lemma 5.1 implies that $g_t(\mathbf{x}) \equiv q_t(\mathbf{x})$ and finishes the proof of the corollary.

□

6 Proof of the Main Theorem

In this section, we give the proof of our main lemma and our main theorem. In addition, we state the consequences of the main theorem for both small formula size and depth of circuits computing polynomials with small bounded degree.

Lemma 6.1 (Main Lemma)**.** *Let $P(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be such that $\deg_y(P) = r$, and also $\deg_{x_i}(P) \le r, \forall i \in \{1, \ldots, n\}$. Let $P'(\mathbf{x}, y) \equiv \frac{\partial P}{\partial y}(\mathbf{x}, y)$. In addition, let $f(\mathbf{x}, y) \in \mathbb{F}(\mathbf{x})[y]$ be in monic standard form and assume it is irreducible over $\mathbb{F}(\mathbf{x})[y]$, satisfying the following conditions:*

(i) $f(\mathbf{x}, y) \mid P(\mathbf{x}, y)$[3],

(ii) $f(\mathbf{0}, y)$ has exactly $\deg_y(f)$ distinct roots[4],

(iii) $P'(\mathbf{0}, y)$ and $f(\mathbf{0}, y)$ share no common roots.

If there exists a formula (circuit) Γ computing P with output gate being an addition gate, $|\Gamma| = s$ and $\mathrm{depth}(\Gamma) = d$, then for every $m \ge 1$, there exist formulas (circuits) Ψ_m and $\widetilde{\Psi}_m$ with each output gate being a multiplication gate, of size

$$\max(|\Psi_m|, |\widetilde{\Psi}_m|) \le 300m^2r^3 \cdot \binom{m+r+1}{r+1} \cdot s$$

and depth $\max(\mathrm{depth}(\Psi_m), \mathrm{depth}(\widetilde{\Psi}_m)) \le d + 3$ such that

$$H^{\mathbf{x}}_{\le m}[\Psi_m] \equiv H^{\mathbf{x}}_{\le m}[\psi_{f,m}(\mathbf{x}, y)] \quad \text{and}$$

$$H^{\mathbf{x}}_{\le m}[\widetilde{\Psi}_m] \equiv H^{\mathbf{x}}_{\le m}[\psi_{\tilde{f},m}(\mathbf{x}, y)].$$

If we require the in-degree of the formula (circuit) to be 2, then the size of Ψ_m or $\widetilde{\Psi}_m$ does not change, and $\max(\mathrm{depth}(\Psi_m), \mathrm{depth}(\widetilde{\Psi}_m)) \le d + 10r \log m$.

[3]Since $P(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$, this condition is equivalent to the existence of $Q(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ such that $f(\mathbf{x}, y) \cdot Q(\mathbf{x}, y) \equiv P(\mathbf{x}, y)$.

[4]Note that we can evaluate $f(\mathbf{x}, y)$ at $\mathbf{x} = \mathbf{0}$, since $f(\mathbf{x}, y)$ is in standard form.

Proof. Let

$$P(\mathbf{x},y) \equiv \sum_{i=0}^{r} C_i(\mathbf{x})y^i \quad \text{and}$$

$$f(\mathbf{x},y) \equiv y^k + \frac{1}{1-g(\mathbf{x})} \cdot \sum_{i=0}^{k-1} f_i(\mathbf{x})y^i \quad \text{s.t. } \gcd(1-g(\mathbf{x}), f_i(\mathbf{x})) = 1.$$

Thus, if $f(\mathbf{0},y) = \prod_{i=1}^{k}(y-\mu_i)$, we have that $P(\mathbf{0},\mu_i) = 0$ and $P'(\mathbf{0},\mu_i) \neq 0$ for all $i \in [k]$. Thus, Lemma 5.1 implies that for each root μ_i of $f(\mathbf{0},y)$, there exists a unique polynomial $q_{i,m}(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ such that $q_{i,m}(\mathbf{0}) = \mu_i$, $\deg(q_{i,m}) \leq m$ and the following polynomial identities hold, for $1 \leq i \leq k$:

$$H^{\mathbf{x}}_{\leq m}[P(\mathbf{x}, q_{i,m}(\mathbf{x}))] \equiv 0.$$

From condition (i), we have:

$$\begin{aligned} f(\mathbf{x},y) \mid P(\mathbf{x},y) &\Rightarrow (1-g(\mathbf{x}))f(\mathbf{x},y) \mid P(\mathbf{x},y) \\ &\Rightarrow \exists R(\mathbf{x},y) \in \mathbb{F}[\mathbf{x},y] \text{ s.t. } P(\mathbf{x},y) \equiv (1-g(\mathbf{x}))f(\mathbf{x},y)R(\mathbf{x},y) \\ &\Rightarrow H^{\mathbf{x}}_{\leq m}[P(\mathbf{x},y)] \equiv H^{\mathbf{x}}_{\leq m}\left[(1-g(\mathbf{x}))f(\mathbf{x},y)R(\mathbf{x},y)\right] \\ &\Rightarrow H^{\mathbf{x}}_{\leq m}[P(\mathbf{x},y)] \equiv H^{\mathbf{x}}_{\leq m}\left[(1-g(\mathbf{x}))\psi_{f,m}(\mathbf{x},y)R(\mathbf{x},y)\right]. \end{aligned} \tag{6}$$

Since $g(\mathbf{0}) = 0$, we have $\psi_{f,m}(\mathbf{0},y) \equiv f(\mathbf{0},y)$, for all $m \in \mathbb{N}$. If, in Corollary 5.2, we let

$$h(\mathbf{x},y) \equiv \psi_{f,m}(\mathbf{x},y) \quad \text{and} \quad Q(\mathbf{x},y) \equiv (1-g(\mathbf{x}))R(\mathbf{x},y),$$

the fact that $\psi_{f,m}(\mathbf{0},y) \equiv f(\mathbf{0},y)$ and equation (6) imply that for each root μ_i of $h(\mathbf{0},y) \equiv \psi_{f,m}(\mathbf{0},y)$, the polynomials $q_{i,m}(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ defined above also satisfy the following polynomial identities:

$$H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x}, q_{i,m}(\mathbf{x}))] \equiv 0.$$

Summarizing, the polynomials $q_{i,m}(\mathbf{x})$ satisfy the following identities:

$$H^{\mathbf{x}}_{\leq m}[P(\mathbf{x}, q_{i,m}(\mathbf{x}))] \equiv 0 \quad \text{and} \quad H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x}, q_{i,m}(\mathbf{x}))] \equiv 0. \tag{7}$$

Moreover, given the above formula Γ computing $P(\mathbf{x},y)$, we can construct formulas Φ_i computing $q_{i,m}(\mathbf{x})$ such that

$$|\Phi_i| \leq 200(mr)^2 \binom{m+r+1}{r+1} \cdot s, \quad \text{and}$$

$$\text{depth}(\Phi_i) \leq \text{depth}(\Gamma) + 2 = d + 2.$$

Now that we have the formulas Φ_i computing each of the polynomials $q_{i,m}(\mathbf{x})$, we just need to show how we can compute $H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x},y)]$. The next claim shows how to compute this quantity.

Claim 6.2. *Let $\mu_1, \ldots, \mu_k$ be the (distinct) roots of $f(\mathbf{0}, y)$ and let $q_{i,m}(\mathbf{x})$ be the corresponding polynomials defined as above. Then, the following holds:*

$$H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x}, y)] \equiv H^{\mathbf{x}}_{\leq m}\left[\prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x}))\right]$$

Proof. To prove the claim, it is enough to prove the following by induction on t: for $1 \leq t \leq k = \deg_y(f)$, there exists polynomials $F_t(\mathbf{x}, y), G_t(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ with $\deg_y(F_t) = k - t$ and $\mathrm{mindeg}_{\mathbf{x}}(G_t(\mathbf{x}, y)) > m$ such that

$$\psi_{f,m}(\mathbf{x}, y) \equiv F_t(\mathbf{x}, y) \cdot \prod_{i=1}^{t}(y - q_{i,m}(\mathbf{x})) + G_t(\mathbf{x}, y).$$

Notice that the fact above is true for $t = 1$, since $y - q_{1,m}(\mathbf{x})$ is monic in y and thereby we can apply the regular univariate division algorithm (regarding the polynomials as polynomials in $\mathbb{F}[\mathbf{x}][y]$) to obtain

$$\psi_{f,m}(\mathbf{x}, y) \equiv F_1(\mathbf{x}, y) \cdot (y - q_{1,m}(\mathbf{x})) + \psi_{f,m}(\mathbf{x}, q_{1,m}(\mathbf{x})).$$

Where $F_1(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$, $\deg_y(F_1) = k - 1$ and we have that $\mathrm{mindeg}_{\mathbf{x}}(\psi_{f,m}(\mathbf{x}, q_{1,m}(\mathbf{x}))) > m$, by equation (7).

Now, suppose that the fact above is true for $t - 1$, where $2 \leq t \leq k$. Since $y - q_{t,m}(\mathbf{x})$ is monic in y, by dividing $F_{t-1}(\mathbf{x}, y)$ by $y - q_{t,m}(\mathbf{x})$, we have:

$$F_{t-1}(\mathbf{x}, y) \equiv F_t(\mathbf{x}, y) \cdot (y - q_{t,m}(\mathbf{x})) + F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x})).$$

Hence, we obtain that

$$\psi_{f,m}(\mathbf{x}, y) \equiv F_t(\mathbf{x}, y) \cdot \prod_{i=1}^{t}(y - q_{i,m}(\mathbf{x})) + F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x})) \cdot \prod_{i=1}^{t-1}(y - q_{i,m}(\mathbf{x})) + G_{t-1}(\mathbf{x}, y).$$

From this equation, it follows that

$$\psi_{f,m}(\mathbf{x}, q_{t,m}(\mathbf{x})) \equiv 0 + F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x})) \cdot \prod_{i=1}^{t-1}(q_{t,m}(\mathbf{x}) - q_{i,m}(\mathbf{x})) + G_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x}))$$

$$\iff F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x})) \cdot \prod_{i=1}^{t-1}(q_{t,m}(\mathbf{x}) - q_{i,m}(\mathbf{x})) \equiv \psi_{f,m}(\mathbf{x}, q_{t,m}(\mathbf{x})) - G_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x})). \quad (8)$$

Since the roots μ_i are all distinct, we have that $\mu_i \neq \mu_j$, for all $i \neq j$ and therefore

$$\mathrm{mindeg}_{\mathbf{x}}\left(\prod_{i=1}^{t-1}(q_{t,m}(\mathbf{x}) - q_{i,m}(\mathbf{x}))\right) = 0,$$

which implies

$$\mathrm{mindeg}_{\mathbf{x}}\left(F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x})) \cdot \prod_{i=1}^{t-1}(q_{t,m}(\mathbf{x}) - q_{i,m}(\mathbf{x}))\right) = \mathrm{mindeg}_{\mathbf{x}}(F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x}))).$$

Therefore, we have

$$\begin{aligned}
\mathrm{mindeg}_{\mathbf{x}}(F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x}))) &= \mathrm{mindeg}_{\mathbf{x}}\left(F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x})) \cdot \prod_{i=1}^{t-1}(q_{t,m}(\mathbf{x}) - q_{i,m}(\mathbf{x}))\right) && \text{(by the above equation)} \\
&= \mathrm{mindeg}_{\mathbf{x}}(\psi_{f,m}(\mathbf{x}, q_{t,m}(\mathbf{x})) - G_{t-1}(\mathbf{x}, q_{it,m}(\mathbf{x}))) && \text{by equation (8)} \\
&> m,
\end{aligned}$$

where the last inequality is true by equation (7) and induction hypothesis.

Hence, by defining

$$G_t(\mathbf{x}, y) = F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x})) \cdot \prod_{i=1}^{t-1}(y - q_{i,m}(\mathbf{x})) + G_{t-1}(\mathbf{x}, y)$$

we have

$$\mathrm{mindeg}_{\mathbf{x}}(G_t(\mathbf{x}, y)) \geq \min\left(\mathrm{mindeg}_{\mathbf{x}}(F_{t-1}(\mathbf{x}, q_{t,m}(\mathbf{x}))), \mathrm{mindeg}_{\mathbf{x}}(G_{t-1}(\mathbf{x}, y))\right) > m.$$

This finishes the inductive argument.

Since

$$\psi_{f,m}(\mathbf{x}, y) \equiv F_k(\mathbf{x}, y) \cdot \prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x})) + G_k(\mathbf{x}, y),$$

and $H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x}, y)]$ is a monic polynomial in y, for $f(\mathbf{x}, y)$ is in monic standard form, we have that $H^{\mathbf{x}}_{\leq m}\left[F_k(\mathbf{x}, y) \cdot \prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x})) + G_k(\mathbf{x}, y)\right]$ is also a monic polynomial in y and therefore

$$H^{\mathbf{x}}_{\leq m}\left[F_k(\mathbf{x}, y) \cdot \prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x})) + G_k(\mathbf{x}, y)\right] \equiv H^{\mathbf{x}}_{\leq m}\left[\prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x})) + G_k(\mathbf{x}, y)\right].$$

Hence, by the induction hypothesis on $G_k(\mathbf{x}, y)$, we have that

$$H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x}, y)] \equiv H^{\mathbf{x}}_{\leq m}\left[\prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x})) + G_k(\mathbf{x}, y)\right] \equiv H^{\mathbf{x}}_{\leq m}\left[\prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x}))\right].$$

$\square$

By this last claim,

$$H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x},y)] \equiv H^{\mathbf{x}}_{\leq m}\left[\prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x}))\right], \tag{9}$$

which implies that $\prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x}))$ is a polynomial that agrees with $\psi_{f,m}(\mathbf{x},y)$ on the homogeneous parts of degree $\leq m$. Hence, the formula Ψ_m given by

$$\Psi_m = \prod_{i=1}^{k}(y - \Phi_i)$$

is such that

$$H^{\mathbf{x}}_{\leq m}[\Psi_m(\mathbf{x},y)] \equiv H^{\mathbf{x}}_{\leq m}\left[\prod_{i=1}^{k}(y - q_{i,m}(\mathbf{x}))\right] \equiv H^{\mathbf{x}}_{\leq m}[\psi_{f,m}(\mathbf{x},y)].$$

Since each Φ_i is a formula such that $\text{depth}(\Phi_i) \leq d+2$, their output gates are addition gates and

$$|\Phi_i| \leq 200(mr)^2\binom{m+r+1}{r+1}\cdot s,$$

we have that

$$|\Psi_m| \leq k\cdot\left(200(mr)^2\binom{m+r+1}{r+1}\cdot s + 2k\right) \leq 300m^2r^3\cdot\binom{m+r+1}{r+1}\cdot s.$$

Since Ψ_m is the product of formulas of depth $\leq d+2$, then we have that $\text{depth}(\Psi_m) \leq d+3$.

By Observation 4.3 and equation (9),

$$H^{\mathbf{x}}_{\leq m}[\psi_{\tilde{f},m}(\mathbf{x},y)] \equiv H^{\mathbf{x}}_{\leq m}\left[\widetilde{\psi}_{f,m}(\mathbf{x},y)\right] \equiv H^{\mathbf{x}}_{\leq m}\left[\prod_{i=1}^{k}(1 - yq_{i,m}(\mathbf{x}))\right],$$

which implies that $\prod_{i=1}^{k}(1 - yq_{i,m}(\mathbf{x}))$ is a polynomial that agrees with $\psi_{\tilde{f},m}(\mathbf{x},y)$ on the homogeneous parts of degree $\leq m$. Hence, the formula $\widetilde{\Psi}_m$ given by

$$\widetilde{\Psi}_m = \prod_{i=1}^{k}(1 - y\Phi_i)$$

is such that $H^{\mathbf{x}}_{\leq m}[\widetilde{\Psi}_m] \equiv H^{\mathbf{x}}_{\leq m}[\psi_{\tilde{f},m}(\mathbf{x},y)]$. It is easy to see that the same size and depth bounds for Ψ_m apply for the formula $\widetilde{\Psi}_m$. In addition, it is clear from the proof that if we restrict the formulas to have in-degree bounded by 2, we obtain the desired depth. □

Theorem 6.3 (Main Theorem)**.** *Let $P(\mathbf{x}) \in \mathbb{F}[\mathbf{x}] \setminus \{0\}$ be such that $\deg_{x_i}(P) \leq r$, $1 \leq i \leq n$, $P(\mathbf{0}) \neq 0$ and let Γ be a formula (circuit) of size s and depth d computing P. Let $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a factor of $P(\mathbf{x})$, and let m be a positive integer. There exists a polynomial $G(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ of* *total degree* $\deg(G) \le 4r^3n^3$ *such that if* $\mathbf{c} \in \mathbb{F}^n$ *satisfies* $G(\mathbf{c}) \neq 0$ *then there exists a formula* Φ_m *whose output gate is a multiplication gate and for which*

$$\mathrm{depth}(\Phi_m) \le d + 4^{5},$$

$$|\Phi_m| \le 60000 m^2 r^8 n \cdot \binom{m+r+1}{r+1} s \quad \textit{and}$$

$$H^{\mathbf{x}}_{\le m}[\Phi_m(\mathbf{x})] \equiv H^{\mathbf{x}}_{\le m}[f(\mathbf{x}+\mathbf{c})].$$

If we require the in-degree of the formula (circuit) to be 2*, then the size of* Φ_m *does not change, and* $\mathrm{depth}(\Phi_m) \le d + 20r\log m$*.*

Proof. The proof of the theorem is by induction on the number of variables. The bound is trivial in the univariate case, since if $f(x), P(x) \in \mathbb{F}[x]$, where $\deg(f) = k \le r$ and $f \mid P$, then we can write

$$f(x) = c \cdot \prod_{i=1}^{k} (x - \mu_i),$$

which can be trivially computed by a formula Ψ of size $\le 50k$ and depth 2. In this case, setting $G(x)$ to be any constant polynomial, for instance $G(x) \equiv 1$, and $\Phi_m = \Psi$, takes care of the base case.

Hence, let's assume that the claim is true for polynomials $P(\mathbf{x}) \in \mathbb{F}[\mathbf{x}] = \mathbb{F}[x_1, \ldots, x_n]$ with $P(\mathbf{0}) \neq 0$, for some $n \ge 1$. Now we will prove that the same bounds hold for polynomials $P(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ s.t. $P(\mathbf{0}, 0) \neq 0$. Let $P(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be a polynomial computed by Γ and $f(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ be a factor of $P(\mathbf{x}, y)$. We can assume that $f(\mathbf{x}, y)$ and $P(\mathbf{x}, y)$ depend on y, otherwise we can simply restrict the circuit Γ to $\Gamma|_{y=0}$, and by the induction hypothesis the result follows.

Let

$$P(\mathbf{x}, y) \equiv \sum_{i=0}^{r} C_i(\mathbf{x}) y^i \quad \text{and}$$

$$f(\mathbf{x}, y) \equiv q(\mathbf{x}) \cdot \prod_{i=1}^{t} f_i(\mathbf{x}, y)^{e_i}, \quad \text{with}$$

$$f_i(\mathbf{x}, y) \equiv \sum_{j=0}^{k_i} f_{ij}(\mathbf{x}) y^j, \text{ where } f_{i0}(\mathbf{x}) \cdot f_{ik_i}(\mathbf{x}) \not\equiv 0, \quad \forall 1 \le i \le t.$$

where each $f_i(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ is an irreducible polynomial. Since $P(\mathbf{0}, 0) \neq 0$, we have that $C_0(\mathbf{x}) \equiv P(\mathbf{x}, 0) \not\equiv 0$, and moreover, that $C_0(\mathbf{0}) \neq 0$. Let

$$u(\mathbf{x}) \equiv f(\mathbf{x}, 0) \equiv q(\mathbf{x}) \cdot \prod_{i=1}^{t} f_{i0}(\mathbf{x})^{e_i}.$$

[5] If the bottom gates are addition gates, then the depth is bounded by $d + 3$.

Notice that $f(\mathbf{x}, y) \mid P(\mathbf{x}, y) \Rightarrow u(\mathbf{x}) \mid C_0(\mathbf{x})$. In addition, notice that $C_0(\mathbf{0}) \neq 0$ and $C_0(\mathbf{x})$ can be computed by the formula $\Gamma|_{y=0}$, where $|\Gamma|_{y=0}| \leq |\Gamma|$ and $\text{depth}(\Gamma|_{y=0}) \leq \text{depth}(\Gamma)$. Therefore, by induction hypothesis, there exists $H(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ with $\deg(H) \leq 4r^3n^3$ such that for any $\mathbf{a} \in \mathbb{F}^n$ for which $H(\mathbf{a}) \neq 0$, there exists a formula Λ_m with output gate being a multiplication gate, such that

$$\text{depth}(\Lambda_m) \leq d + 4,$$

$$|\Lambda_m| \leq 60000m^2r^8n \cdot \binom{m+r+1}{r+1} s \quad \text{and}$$

$$H^{\mathbf{x}}_{\leq m}[\Lambda_m(\mathbf{x})] \equiv H^{\mathbf{x}}_{\leq m}[u(\mathbf{x} + \mathbf{a})].$$

Now that we have an approximation to the factor $u(\mathbf{x})$, which is the constant term of the polynomial $f(\mathbf{x}, y)$ when seen as a polynomial in the variable y, we want to use Lemma 6.1 to find the factors of $f(\mathbf{x}, y)$ that contain y. For this, we will first need to find polynomials $D_i(\mathbf{x}, y)$ with small formulas such that $f_i(\mathbf{x}, y) \mid D_i(\mathbf{x}, y)$ and each D_i is square-free with respect to $f_i(\mathbf{x}, y)$.

Fortunately, Lemma 2.6 tells us that for each (irreducible) polynomial $f_i(\mathbf{x}, y)$, we can find formulas Δ_i of size $\leq 9r^2|\Gamma|$ computing polynomials $D_i(\mathbf{x}, y)$ such that $\deg_{x_j}(D_i) \leq r$, $1 \leq j \leq n$, $\deg_y(D_i) \leq r$, $f_i(\mathbf{x}, y) \mid D_i(\mathbf{x}, y)$ but $f_i(\mathbf{x}, y) \nmid \dfrac{\partial D_i}{\partial y}(\mathbf{x}, y)$. Moreover these formulas have an addition gate as output gate.

Since $f_i(\mathbf{x}, y)$ is irreducible with $\deg_y(f_i) \geq 1$ and $f_i(\mathbf{x}, y) \nmid \dfrac{\partial D_i}{\partial y}(\mathbf{x}, y)$, Lemma 3.6 implies that there exists a polynomial $G_i(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ with

$$\deg(G_i) \leq 2\deg(f_i)^2 + 2\deg(f_i)\deg(D_i) \leq 4r^2n^2$$

such that for any $\mathbf{c} \in \mathbb{F}^n$ where $G_i(\mathbf{c}) \neq 0$ we have that $\mathbf{c}$ properly splits $f_i(\mathbf{c}, y)$ with respect to $\dfrac{\partial D_i}{\partial y}(\mathbf{c}, y)$.

Let

$$G(\mathbf{x}, y) \equiv H(\mathbf{x}) \cdot C_0(\mathbf{x}) \cdot \prod_{i=1}^{t} G_i(\mathbf{x}) \quad \text{and} \quad (\mathbf{c}, \gamma) \in \mathbb{F}^{n+1} \quad \text{be s.t.} \quad G(\mathbf{c}, \gamma) \neq 0.\text{[6]}$$

[6] At first, it may seem strange that $G(\mathbf{x}, y)$ does not depend on the variable y, since if we continued this argument by induction we would arrive at the conclusion that $G(\mathbf{x}, y)$ is the constant polynomial. However, notice that even though $H(\mathbf{x})$ does not depend on the variable x_n, the polynomial $G(\mathbf{x}, y)$ depends on x_n, since the polynomials $C_0(\mathbf{x})$ and $G_i(\mathbf{x})$ depend on x_n. The right way to see this dependence is the following: $G(\mathbf{x}, y)$ depends on every variable except the variable used by the lifting procedure, which in this case is the variable y. Hence, we will have that $H(\mathbf{x})$ depends on all the variables except x_n (if we choose to perform the lifting with respect to x_n).

Denote

$$Q(\mathbf{x},y) \equiv P(\mathbf{x}+\mathbf{c},y) \equiv \sum_{i=0}^{r} Q_i(\mathbf{x})y^i,$$

$$h_i(\mathbf{x},y) \equiv f_i(\mathbf{x}+\mathbf{c},y) \equiv \sum_{j=0}^{k_i} h_{ij}(\mathbf{x})y^j \quad \text{and}$$

$$h(\mathbf{x},y) \equiv f(\mathbf{x}+\mathbf{c},y) \equiv q(\mathbf{x}+\mathbf{c}) \cdot \prod_{i=1}^{t} h_i(\mathbf{x},y)^{e_i}.$$

Since $h_{i0}(\mathbf{x}) \equiv f_{i0}(\mathbf{x}+\mathbf{c},0) \mid P(\mathbf{x}+\mathbf{c},0) \equiv C_0(\mathbf{x}+\mathbf{c})$ and $C_0(\mathbf{c}) \neq 0$ (because $G(\mathbf{c},\gamma) \neq 0$), we have that $h_{i0}(\mathbf{0}) \neq 0$, for all $1 \leq i \leq t$. Hence, after normalization by a proper field element, we can write each h_{i0} in the following form:

$$h_{i0}(\mathbf{x}) = 1 - g_i(\mathbf{x}), \quad \text{where } g_i(\mathbf{0}) \equiv 0.$$

In addition, notice that $f_{ik_i}(\mathbf{x}) \not\equiv 0 \Rightarrow h_{ik_i}(\mathbf{x}) \equiv f_{ik_i}(\mathbf{x}+\mathbf{c}) \not\equiv 0$.

Moreover, notice that $f_i(\mathbf{x},y)$ is irreducible with $f_{i0}(\mathbf{x}) \cdot f_{ik_i}(\mathbf{x}) \not\equiv 0$ implies that $h_i(\mathbf{x},y)$ is irreducible with $h_{i0}(\mathbf{x}) \cdot h_{ik_i}(\mathbf{x}) \not\equiv 0$, which implies (by Corollary 2.10) that the polynomial $\widetilde{h}_i(\mathbf{x},y) \equiv \sum_{j=0}^{k_i} h_{ij}(\mathbf{x})y^{k_i-j}$ is irreducible in $\mathbb{F}[\mathbf{x},y]$. Hence, we have that $\ell_i(\mathbf{x},y) \equiv \dfrac{\widetilde{h}_i(\mathbf{x},y)}{h_{i0}(\mathbf{x})}$ is a monic irreducible standard form in $\mathbb{F}(\mathbf{x})[y]$.

Because $f_i(\mathbf{x},y) \mid D_i(\mathbf{x},y)$ and $f_i(\mathbf{x},y) \nmid \dfrac{\partial D_i}{\partial y}(\mathbf{x},y)$, by Lemma 2.9 we obtain that $h_i(\mathbf{x},y) \mid E_i(\mathbf{x},y) \equiv D_i(\mathbf{x}+\mathbf{c},y)$ and $h_i(\mathbf{x},y) \nmid \dfrac{\partial E_i}{\partial y}(\mathbf{x},y) \equiv \dfrac{\partial D_i}{\partial y}(\mathbf{x}+\mathbf{c},y)$.

Since $h_i(\mathbf{0},y) \equiv f_i(\mathbf{c},y)$, we also have that $h_i(\mathbf{0},y)$ has no common roots with $\dfrac{\partial E_i}{\partial y}(\mathbf{0},y)$. The following claim shows that $\ell_i(\mathbf{x},y)$ satisfies the conditions of Lemma 6.1.

Claim 6.4. *For each $i \in \{1,\ldots,t\}$, the monic irreducible standard form $\ell_i(\mathbf{x},y) \equiv \dfrac{\widetilde{h}_i(\mathbf{x},y)}{h_{i0}(\mathbf{x})}$ and the polynomial $\tilde{E}_i(\mathbf{x},y)$ satisfy the conditions of Lemma 6.1.*

Proof of claim. Notice that conditions (i) and (ii) from Lemma 6.1 follow from the fact that $h_i(\mathbf{x},y) \mid E_i(\mathbf{x},y)$ and Lemmas 2.9 and 3.6. Condition (iii) follows from the fact that $\dfrac{h_i(\mathbf{0},y)}{h_{i0}(\mathbf{0})} \equiv h_i(\mathbf{0},y)$ shares no common roots with $\dfrac{\partial E_i}{\partial y}(\mathbf{0},y)$ and from Lemma 2.11.

This finishes the proof of the claim. $\square$

Now that we have rational functions in monic standard form that are, in a certain sense, computing the reversal of each $f_i(\mathbf{x},y)$, we can use the main lemma to lift the factorization of the approximation polynomial of $f_i(\mathbf{x},y)/f_{i0}(\mathbf{x})$.[7]

Since each $\ell_i(\mathbf{x},y)$ and $\tilde{E}_i(\mathbf{x},y)$ satisfy the conditions of Lemma 6.1, and $\tilde{E}_i(\mathbf{x},y)$ can be computed by a formula Υ_i of size $|\Upsilon_i| \leq 180r^4 \cdot |\Gamma| = 180r^4 s$ and depth $\mathrm{depth}(\Upsilon_i) \leq d+1$ (since Υ_i is the shift of $\widetilde{\Delta}_i$), we have that there exists a formula $\Psi_{i,m}$ having as output gate a multiplication gate, $\mathrm{depth}(\Psi_{i,m}) \leq \mathrm{depth}(\Upsilon_i) + 3 \leq d+4$ and size

$$|\Psi_{i,m}| \leq 300m^2r^3 \cdot \binom{m+r+1}{r+1} \cdot 180r^4 \cdot s \leq 60000m^2r^7 \cdot \binom{m+r+1}{r+1} \cdot s$$

such that

$$H^{\mathbf{x}}_{\leq m}[\Psi_{i,m}] \equiv H^{\mathbf{x}}_{\leq m}[\psi_{\widetilde{\ell}_i(\mathbf{x},y),m}(\mathbf{x},y)].$$

By Observation 4.3, we have that

$$H^{\mathbf{x}}_{\leq m}[h_{i0}(\mathbf{x}) \cdot \psi_{\widetilde{\ell}_i(\mathbf{x},y),m}(\mathbf{x},y)] \equiv H^{\mathbf{x}}_{\leq m}[\widetilde{\ell}_i(\mathbf{x},y) \cdot h_{i0}(\mathbf{x})] \equiv H^{\mathbf{x}}_{\leq m}[h_i(\mathbf{x},y)], \quad \text{and also}$$

$$H^{\mathbf{x}}_{\leq m}[h_{i0}(\mathbf{x}) \cdot \psi_{\widetilde{\ell}_i(\mathbf{x},y),m}(\mathbf{x},y+\gamma)] \equiv H^{\mathbf{x}}_{\leq m}[h_i(\mathbf{x},y+\gamma)].$$

In addition, from the formulas $\Psi_{i,m}$ and from the fact that $\sum_{i=1}^{t} e_i \leq r$, we have that the formula given by $\Psi_m = \prod_{i=1}^{t} \Psi_{i,m}^{e_i}$ is of size

$$|\Psi_m| \leq \sum_{i=1}^{t} e_i \cdot |\Psi_{i,m}| \leq r \cdot \max_{1\leq i\leq t}(|\Psi_{i,m}|) \leq 60000m^2r^8 \cdot \binom{m+r+1}{r+1} \cdot s$$

and computes the following polynomial:

$$H^{\mathbf{x}}_{\leq m}[\Psi_m(\mathbf{x},y)] \equiv H^{\mathbf{x}}_{\leq m}\left[\prod_{i=1}^{t} \psi_{\widetilde{\ell}_i(\mathbf{x},y),m}(\mathbf{x},y+\gamma)^{e_i}\right].$$

Now that we found a formula computing the approximation polynomials $\psi_{\widetilde{\ell}_i(\mathbf{x},y),m}(\mathbf{x},y+\gamma)$, we can multiply them by $h_{i0}(\mathbf{x},y)$ and via Observation 4.3 obtain the polynomials $h_i(\mathbf{x},y)$, which are the shifts of $f_i(\mathbf{x},y)$. Since Ψ_m computes all of the approximation polynomials, and Λ_m computes all of the leading coefficients, by combining them we can recover the factor $f(\mathbf{x},y)$. This is what we do next.

Multiplying Ψ_m by Λ_m, we have that the formula $\Phi_m = \Lambda_m \cdot \Psi_m$ is such that

$$|\Phi_m| \leq |\Lambda_m| + |\Psi_m| \leq 60000m^2r^8(n+1) \cdot \binom{m+r+1}{r+1} \cdot s$$

[7] In actuality, we are performing a lift of a shift of $f_i(\mathbf{x},y)$.

and

$$
\begin{aligned}
H^{\mathbf{x}}_{\leq m}[\Phi_m(\mathbf{x},y)] &\equiv H^{\mathbf{x}}_{\leq m}[\Lambda_m \cdot \Psi_m] \\
&\equiv H^{\mathbf{x}}_{\leq m}\left[u(\mathbf{x}+\mathbf{c}) \cdot \prod_{i=1}^{t} \psi_{\frac{h_i(\mathbf{x},y)}{h_{i0}(\mathbf{x})},m}(\mathbf{x}, y+\gamma)^{e_i}\right] \\
&\equiv H^{\mathbf{x}}_{\leq m}\left[q(\mathbf{x}+\mathbf{c}) \cdot \prod_{i=1}^{t} f_{i0}(\mathbf{x}+\mathbf{c})^{e_i} \cdot \prod_{i=1}^{t} \psi_{\frac{h_i(\mathbf{x},y)}{h_{i0}(\mathbf{x})},m}(\mathbf{x}, y+\gamma)^{e_i}\right] \\
&\equiv H^{\mathbf{x}}_{\leq m}\left[q(\mathbf{x}+\mathbf{c}) \cdot \prod_{i=1}^{t} \left(h_{i0}(\mathbf{x}) \cdot \psi_{\frac{h_i(\mathbf{x},y)}{h_{i0}(\mathbf{x})},m}(\mathbf{x}, y+\gamma)\right)^{e_i}\right] \\
&\equiv H^{\mathbf{x}}_{\leq m}\left[q(\mathbf{x}+\mathbf{c}) \cdot \prod_{i=1}^{t} h_i(\mathbf{x}, y+\gamma)^{e_i}\right] \\
&\equiv H^{\mathbf{x}}_{\leq m}\left[q(\mathbf{x}+\mathbf{c}) \cdot \prod_{i=1}^{t} f_i(\mathbf{x}+\mathbf{c}, y+\gamma)^{e_i}\right] \\
&\equiv H^{\mathbf{x}}_{\leq m}[f(\mathbf{x}+\mathbf{c}, y+\gamma)].
\end{aligned}
$$

Since

$$\deg(G(\mathbf{x},y)) \leq \deg(H) + \deg(C_0) + \sum_{i=1}^{t} \deg(G_i) \leq 4r^3n^3 + rn + r \cdot 4r^2n^2 \leq 4r^3(n+1)^3,$$

this finishes the induction, and therefore the proof of the theorem. It is clear from the proof, via Observation 2.7, that if we restrict the in-degree to 2, we obtain the desired bound on the depth. □

As a corollary of the main theorem, we obtain:

Corollary 6.5 (Small Formula – Restatement of Theorem 1)**.** *Let $P(\mathbf{x}) \in \mathbb{F}[\mathbf{x}] \setminus \{0\}$ be such that $\deg_{x_i}(P) \leq r$, $1 \leq i \leq n$, and let $f(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ be a factor of P. If there exists a formula Γ of size s and depth d computing P, then there exists a formula Φ of depth $\mathrm{depth}(\Phi) \leq d+5$ and size*

$$|\Phi| = O\left(n^3r^{12} \cdot \binom{nr+r+1}{r+1} s\right) = \mathsf{poly}((nr)^r, s)$$

such that

$$\Phi(\mathbf{x}) \equiv f(\mathbf{x}).$$

If we require the in-degree of the formula (circuit) to be 2, then the size of Φ does not change, and $\mathrm{depth}(\Phi) \leq d + 30r\log(nr)$.

Proof. Let $\mathbf{c} \in \mathbb{F}^n$ be such that $P(\mathbf{c}) \neq 0$, such a $\mathbf{c}$ exists since $P(\mathbf{x})$ is nonzero. This implies that $Q(\mathbf{x}) \equiv P(\mathbf{x}+\mathbf{c})$ is computed by the formula $\Delta(\mathbf{x}) = \Gamma(\mathbf{x}+\mathbf{c})$, of size $\leq 2|\Gamma| = 2s$, depth $\mathrm{depth}(\Delta) \leq d+1$ and is such that $Q(\mathbf{0}) = P(\mathbf{c}) \neq 0$. Hence, by Theorem 6.3, we have that there exists polynomial $G(\mathbf{x}) \in \mathbb{F}[\mathbf{x}]$ of degree $\deg(G) \leq 4r^3n^3$ such that for any $\mathbf{a} \in \mathbb{F}^n$ for which $G(\mathbf{a}) \neq 0$, there is a formula Φ_{nr} whose output gate is a multiplication gate for which $\mathrm{depth}(\Phi_{nr}) \leq \mathrm{depth}(\Delta) + 3 \leq d + 4$, of size

$$|\Phi_{nr}| \leq 120000(nr)^2 r^8 n \cdot \binom{nr+r+1}{r+1} s \text{ and such that}$$

$$H^{\mathbf{x}}_{\leq nr}[\Phi_{nr}(\mathbf{x})] \equiv H^{\mathbf{x}}_{\leq nr}[f(\mathbf{x}+\mathbf{c}+\mathbf{a})] \equiv f(\mathbf{x}+\mathbf{c}+\mathbf{a}), \quad \text{since } nr \geq \deg(f).$$

By the interpolation Lemma 2.4, we obtain that there exists a formula Φ' of size

$$|\Phi'| \leq 9r^2 \cdot |\Phi_{nr}|$$

and depth $\mathrm{depth}(\Phi') \leq d + 5$ such that $\Phi'(\mathbf{x}) \equiv f(\mathbf{x}+\mathbf{c}+\mathbf{a})$. By shifting the inputs of the formula Φ' by $-(\mathbf{a}+\mathbf{c})$, we have that the new formula just obtained, call it Φ, computes the polynomial $f(\mathbf{x})$, as we wanted. It is easy to see that Φ has the desired upper bound on its size. It is also clear from the proof that if we restrict the in-degree of the formulas (circuits) to be 2, we obtain the desired bounds on the depth. This finishes the proof. $\square$

7 Conclusion

Besides solving a question posed by Kopparty et al. [KSS14] and Open Problem 19 in [SY10] for the class of polynomials of bounded individual degree, notice that Lemma 6.1 and Theorem 6.3 also provide a framework to obtain formulas (circuits) for the approximate roots of a polynomial into actual formulas (circuits) for factors of the same polynomial. Since Lemma 6.1, and therefore Theorem 6.3, uses the Approximation Lemma (Lemma 5.1) as a black-box, any improvements on Lemma 5.1 would lead to better bounds on the size of the formulas for the factors of the input polynomial. Hence, if one can remove the exponential dependence on the parameter r (the bound on the individual degrees) in the Approximation Lemma, one can fully solve the questions above. This is the main open question left by this work.

Acknowledgments

The author would like to thank his advisor Zeev Dvir for all the helpful discussions and encouragement throughout the course of this work.

References

[CLO06] D. A. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 2006.

[CR88] Benny Chor and Ronald L. Rivest. A knapsack-type public key cryptosystem based on arithmetic in finite fields. *IEEE Transactions on Information Theory*, 34(5):901–909, 1988.

[DSY09] Z. Dvir, A. Shpilka, and A. Yehudayoff. Hardness-randomness tradeoffs for bounded depth arithmetic circuits. *SIAM J. on Computing*, 39(4):1279–1293, 2009.

[GG99] J. von zur Gathen and J. Gerhard. *Modern computer algebra*. Cambridge University Press, 1999.

[GK85] J. Von Zur Gathen and E. Kaltofen. Factoring sparse multivariate polynomials. *Journal of Computer and System Sciences*, 31(2):265–287, 1985.

[GS06] V. Guruswami and M. Sudan. Improved decoding of reed-solomon and algebraic-geometry codes. *IEEE Trans. Inf. Theor.*, 45(6):1757–1767, September 2006.

[Kal85] E. Kaltofen. Polynomial-time reductions from multivariate to bi- and univariate integral polynomial factorization. *SIAM J. on computing*, 14(2):469–489, 1985.

[Kal89] E. Kaltofen. Factorization of polynomials given by straight-line programs. In S. Micali, editor, *Randomness in Computation*, volume 5 of *Advances in Computing Research*, pages 375–412. 1989.

[Kal03] E. Kaltofen. Polynomial factorization: a success story. In *ISSAC*, pages 3–4, 2003.

[KI04] V. Kabanets and R. Impagliazzo. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity*, 13(1-2):1–46, 2004.

[KSS14] Swastik Kopparty, Shubhangi Saraf, and Amir Shpilka. Equivalence of polynomial identity testing and deterministic multivariate polynomial factorization. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 169–180, 2014.

[LJ99] Hendrik W Lenstra Jr. Finding small degree factors of lacunary polynomials. *Number theory in progress*, 1:267–276, 1999.

[LLL82] A. K. Lenstra, H. W. Lenstra, and L. Lovász. Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, 1982.

[Sud97] Madhu Sudan. Decoding of reed solomon codes beyond the error-correction bound. *Journal of Complexity*, 13(1):180 – 193, 1997.

[SV10] A. Shpilka and I. Volkovich. On the relation between polynomial identity testing and finding variable disjoint factors. In *ICALP (1)*, pages 408–419, 2010.

[SY10] A. Shpilka and A. Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3-4):207–388, 2010.

A Proof of Lemmas From Section 2

Proof of Lemma 2.4. Let $a_0, a_1, \ldots, a_r$ be $r+1$ distinct elements of $\mathbb{F}$. For each $i \in \{0, 1, \ldots, r\}$, let

$$\Gamma_i(\mathbf{x}) = \Gamma(\mathbf{x}, a_i)$$

be the restriction of the circuit $\Gamma(\mathbf{x}, y)$ when $y = a_i$. It is clear that $|\Gamma_i| \le |\Gamma|$. Since

$$\begin{pmatrix} 1 & a_0 & a_0^2 & \ldots & a_0^r \\ 1 & a_1 & a_1^2 & \ldots & a_1^r \\ 1 & a_2 & a_2^2 & \ldots & a_2^r \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_r & a_r^2 & \ldots & a_r^r \end{pmatrix} \cdot \begin{pmatrix} P_0(\mathbf{x}) \\ P_1(\mathbf{x}) \\ P_2(\mathbf{x}) \\ \vdots \\ P_r(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} P(\mathbf{x}, a_0) \\ P(\mathbf{x}, a_1) \\ P(\mathbf{x}, a_2) \\ \vdots \\ P(\mathbf{x}, a_r) \end{pmatrix}$$

and the matrix on the left side is a Vandermonde matrix, which is known to be invertible, by left-multiplying by its inverse we obtain:

$$\begin{pmatrix} 1 & a_0 & a_0^2 & \ldots & a_0^r \\ 1 & a_1 & a_1^2 & \ldots & a_1^r \\ 1 & a_2 & a_2^2 & \ldots & a_2^r \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_r & a_r^2 & \ldots & a_r^r \end{pmatrix}^{-1} \cdot \begin{pmatrix} P(\mathbf{x}, a_0) \\ P(\mathbf{x}, a_1) \\ P(\mathbf{x}, a_2) \\ \vdots \\ P(\mathbf{x}, a_r) \end{pmatrix} = \begin{pmatrix} P_0(\mathbf{x}) \\ P_1(\mathbf{x}) \\ P_2(\mathbf{x}) \\ \vdots \\ P_r(\mathbf{x}) \end{pmatrix} \tag{10}$$

Let M be the matrix in equation (10), where the $(i,j)^{th}$ entry of M is denoted by m_{ij}, where the indices of M range from 0 to r. Since circuit Γ_i computes the polynomial $P(\mathbf{x}, a_i)$, by equation (10) we have that

$$P_i(\mathbf{x}) = \sum_{j=0}^{r} m_{ij} \cdot \Gamma_j(\mathbf{x}) \tag{11}$$

Which implies that $P_i(\mathbf{x})$ can be computed by a circuit Φ_i of size $\le r \cdot (|\Gamma| + 2)$, for each Γ_j can be computed by a circuit of size $\le |\Gamma|$ and it takes $\le 2r$ gates to compute the expression in equation (11).

Since we did not reuse any of the restrictions Γ_j neither the gates used to compute expression (11), we have that if Γ is a formula then so will Φ_i be a formula. $\square$

Proof of Lemma 2.6. If the topmost gate of Γ is an addition gate, then for each $1 \le i \le t$, let $d \ge 0$ be the first integer such that $g(\mathbf{x}, y) \mid \dfrac{\partial^d P}{(\partial y)^d}(\mathbf{x}, y)$ and $g(\mathbf{x}, y) \nmid \dfrac{\partial^{d+1} P}{(\partial y)^{d+1}}(\mathbf{x}, y)$. In this case, let $D(\mathbf{x}, y) \equiv \dfrac{\partial^d P}{(\partial y)^d}(\mathbf{x}, y)$. Lemma 2.5 tells us that there exists a formula Δ that computes $D(\mathbf{x}, y)$ and has the required properties.

If the topmost gate of Γ is a product gate, we have that $\Gamma = \displaystyle\prod_{j=1}^{\ell} \Gamma_j$, where each Γ_j is a formula with an output gate being an addition gate, $\text{depth}(\Gamma_j) \le \text{depth}(\Gamma) - 1$ and $|\Gamma_j| < |\Gamma|$.

Let P_j be the polynomial computed by formula Γ_j. In this case, let $k \in \{1, \ldots, \ell\}$ be such that $g(\mathbf{x}, y) \mid P_k(\mathbf{x}, y)$ and $d \geq 0$ be the first integer such that $g(\mathbf{x}, y) \mid \dfrac{\partial^d P_k}{(\partial y)^d}(\mathbf{x}, y)$ and $g(\mathbf{x}, y) \nmid \dfrac{\partial^{d+1} P_k}{(\partial y)^{d+1}}(\mathbf{x}, y)$. In this case, let $D(\mathbf{x}, y) \equiv \dfrac{\partial^d P_k}{(\partial y)^d}(\mathbf{x}, y)$. Again, Lemma 2.5 tells us that there exist a formula Δ that computes $D(\mathbf{x}, y)$ and has the required properties. $\square$

Proof of Lemma 2.8. By Lemma 2.4, from Γ we can obtain circuits Φ_i computing $P_i(\mathbf{x})$ such that $|\Phi_i| \leq 3r \cdot |\Gamma|$. Now, with a formula Λ of size $\leq 2r$ one can compute all powers y^i, where $0 \leq i \leq r$. And finally, with a formula of size $\leq 2r$ (using the output gates of Φ_i and the gates of Λ as inputs), one can compute the expression

$$\sum_{i=0}^{r} y^{r-i} P_i(\mathbf{x}).$$

Hence, we obtain a circuit Δ computing $\tilde{P}$ and the size of Δ is upper bounded by

$$|\Delta| \leq r \cdot (3r \cdot |\Gamma|) + 4r \leq 8r^2 |\Gamma|.$$

Since we did not reuse any gates in this construction, notice that if Γ is a formula then Δ will also be a formula. $\square$

Proof of Lemma 2.9. If $f \mid P$, then there exists $g(\mathbf{x}, y) \in \mathbb{F}[\mathbf{x}, y]$ such that $P(\mathbf{x}, y) \equiv f(\mathbf{x}, y) \cdot g(\mathbf{x}, y)$. If we let $g(\mathbf{x}, y) \equiv \sum_{i=0}^{r-k} g_i(\mathbf{x}) y^i$, then $P(\mathbf{x}, y) \equiv f(\mathbf{x}, y) \cdot g(\mathbf{x}, y)$ implies that

$$P_t(\mathbf{x}) \equiv \sum_{i=0}^{t} f_{t-i}(\mathbf{x}) g_i(\mathbf{x}), \quad \text{for all} \quad 0 \leq t \leq r.$$

Since $P_r(\mathbf{x}) \cdot P_0(\mathbf{x}) \not\equiv 0$ and $f_k(\mathbf{x}) \cdot f_0(\mathbf{x}) \not\equiv 0$, we must have that $g_{r-k}(\mathbf{x}) \cdot g_0(\mathbf{x}) \not\equiv 0$. Therefore, the reversal of g is well defined.

Let $\tilde{g}(\mathbf{x}, y) \equiv \sum_{i=0}^{r-k} g_i(\mathbf{x}) y^{r-k-i}$. Then, notice that

$$\begin{aligned}
\tilde{f}(\mathbf{x}, y) \cdot \tilde{g}(\mathbf{x}, y) &\equiv \left(\sum_{i=0}^{k} f_i(\mathbf{x}) y^{k-i} \right) \cdot \left(\sum_{i=0}^{r-k} g_i(\mathbf{x}) y^{r-k-i} \right) \\
&\equiv \sum_{t=0}^{r} \left(\sum_{i=0}^{t} f_{t-i}(\mathbf{x}) g_i(\mathbf{x}) \right) y^{r-t} \equiv \sum_{t=0}^{r} P_t(\mathbf{x}) y^{r-t} \equiv \tilde{P}(\mathbf{x}, y).
\end{aligned}$$

This last equation implies that $\tilde{f} \mid \tilde{P}$. Analogously, we can show that $\tilde{f} \mid \tilde{P} \Rightarrow f \mid P$. $\square$

Proof of Lemma 2.11. Let $f(x) \equiv \alpha \cdot \prod_{j=1}^{D_f} (x - \lambda_j)^{d_j}$ and $f(x) \equiv \beta \cdot \prod_{j=1}^{D_g} (x - \mu_j)^{e_j}$, where $\alpha, \beta \in \mathbb{F}$. Since f and g share no common roots, it must be the case that $\lambda_i \neq \mu_j$, for all $i \in \{1, \ldots, D_f\}$ and $j \in \{1, \ldots, D_g\}$.

Since

$$\widetilde{f}(x) \equiv \alpha\lambda \cdot \prod_{j=1}^{D_f} \left(x - \frac{1}{\lambda_j} \right)^{d_j}$$

where $\lambda = \prod_{j=1}^{D_f} (-\lambda_j^{d_j})$, and

$$\widetilde{g}(x) \equiv \alpha\mu \cdot \prod_{j=1}^{D_g} \left(x - \frac{1}{\mu_j} \right)^{e_j}$$

where $\lambda = \prod_{j=1}^{D_g} (-\mu_j^{e_j})$ we have that no root of $\widetilde{f}(x)$ is a root of $\widetilde{g}(x)$. $\square$