Extractors and Rank Extractors for Polynomial Sources

Zeev Dvir[*] Ariel Gabizon[†] Avi Wigderson[‡]

Abstract

In this paper we construct explicit deterministic extractors from *polynomial sources*, which are distributions sampled by low degree multivariate polynomials over finite fields. This naturally generalizes previous work on extraction from affine sources (which are degree 1 polynomials). A direct consequence is a deterministic extractor for distributions sampled by polynomial size arithmetic circuits over exponentially large fields. The steps in our extractor construction, and the tools (mainly from algebraic geometry) that we use for them, are of independent interest:

The first step is a construction of *rank extractors*, which are polynomial mappings which "extract" the algebraic rank from any system of low degree polynomials. More precisely, for any n polynomials, k of which are algebraically independent, a rank extractor outputs k algebraically independent polynomials of slightly higher degree. The rank extractors we construct are applicable not only over finite fields but also over fields of characteristic zero.

The next step is relating algebraic independence to min-entropy. We use a theorem of Wooley to show that these parameters are tightly connected. This allows replacing the algebraic assumption on the source (above) by the natural information theoretic one. It also shows that a rank extractor is already a high quality *condenser* for polynomial sources over polynomially large fields.

Finally, to turn the condensers into extractors, we employ a theorem of Bombieri, giving a character sum estimate for polynomials defined over curves. It allows extracting all the randomness (up to a multiplicative constant) from polynomial sources over exponentially large prime fields.

[*] Department of Computer Science, Weizmann institute of science, Rehovot, Israel. zeev.dvir@weizmann.ac.il. Research supported by Binational Science Foundation (BSF) grant and by Minerva Foundation grant.

[†] Department of Computer Science, Weizmann institute of science, Rehovot, Israel. ariel.gabizon@weizmann.ac.il. Research supported by Binational Science Foundation (BSF) grant and by Minerva Foundation grant.

[‡] Institute for Advanced Study, Princeton, NJ, avi@ias.edu

1 Introduction

Randomness extraction has been a major research area for nearly two decades. The functions studied and constructed in this theory: extractors, dispersers, condensers, samplers, etc., turn out to have numerous applications. While they are designed to convert weak sources of randomness into "high quality" random bits, they end up being essential in applications where randomness is not even an issue, such as expander constructions [WZ99], error correction [TSZ01] and metric embedding [Ind07], to name but a few examples.

Most of the aforementioned research has concentrated on the so-called "seeded" extractors, which allow the use of an auxiliary short truly random seed, and enables handling extremely general classes of weak sources. An excellent survey of this broad field is [Sha02]. More recently there has been a burst of activity on "seedless" or "deterministic" extractors, which use no additional random "seed". The general question is for which classes of distributions deterministic extraction is possible. The main types of sources for which progress has been made include the following (somewhat overlapping) classes.

- *Few independent sources*: the given distribution is of several, independent weak sources, as in e.g. [Vaz87, CG88, BIW04, BKS$^+$05, Raz05, Rao06, BRSW06].
- *Computational sources*: the given distribution is the output of some (space- or time-) efficient algorithm on a uniformly random input, as in e.g. [vN51, Blu86, TV00, KRVZ06].
- *Bit-fixing sources*: the given distribution is fixed in some coordinates, and independent in others, as in e.g. [CGH$^+$85, KZ03, GRS04]
- *Affine sources*: the given distribution is the output of some affine map, applied to a random input as in e.g. [BKS$^+$05, Bou07, GR05]

Since our work is best viewed as extending the last class of sources, let us describe these results in some more detail. An *affine source* over a finite field $\mathbb{F}$ is a random variable that is uniformly distributed on some k-dimensional affine subspace of $\mathbb{F}^n$. Such a distribution is usually described by a non-degenerate affine mapping $x(t) : \mathbb{F}^k \to \mathbb{F}^n$ defined by n linear functions

$$x(t) = (x_1(t_1, \dots, t_k), \dots, x_n(t_1, \dots, t_k)),$$

in k variables. The affine source is thought of as the output of $x(t)$ on a uniformly chosen input $t \in \mathbb{F}^k$. Clearly, the entropy (and more importantly, min-entropy) of such sources is $k \cdot \log |\mathbb{F}|$, where all logarithms in this paper are base two. We refer to k as the *rank* of the source and make all asymptotic statements with respect to n.

The works of Barak et al [BKS$^+$05] and of Bourgain [Bou07] deal with the case of the binary field $\mathbb{F}_2$. The first gives an explicit disperser, and the second an extractor, for the case where $k = \Omega(n)$. In particular, Bourgain [Bou07] extracts a constant fraction the entropy with exponentially small error for such k. No explicit construction is known for smaller rank (over $\mathbb{F}_2$) despite the fact that, non explicitly, extractors exist even for logarithmic rank.

Gabizon and Raz [GR05] show that if the field $\mathbb{F}$ is polynomially large, then one can even handle the case of 1-dimensional affine sources (distributions on affine lines). They show how to construct a deterministic extractor that extracts almost all the entropy (with polynomial error) for any given k, for fields $\mathbb{F}$ of size polynomial in n.

1.1 Low Degree Polynomial Sources

A natural generalization of affine sources is allowing sources that arise from low-degree multivariate polynomials. We note that while low-degree polynomials play an essential role in complexity theory, extraction from sources defined by such polynomials has apparently not been studied before.

Let $\mathbb{F}$ be a field (finite or infinite). For integers $k \leq n$ and d we consider the family of all mappings $x : \mathbb{F}^k \to \mathbb{F}^n$ that are defined by polynomials of total degree at most d (we denote our mapping by x since this will represent our source). That is,

$$x(t) = (x_1(t_1, \ldots, t_k), \ldots, x_n(t_1, \ldots, t_k)),$$

where, for each $1 \leq i \leq n$, the coordinate x_i of the mapping is a k-variate polynomial of total degree at most d. We denote this set of mappings by $\mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$. We will focus on the case where the field $\mathbb{F}$ is much larger than d and will specify in each result how large the field has to be.

For affine sources we have the requirement that the affine mapping defining the source is non-degenerate. This ensures that the source sampled by this mapping has 'enough' entropy. We would like to extend this requirement also to the case of low degree mappings in $\mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$. The way to generalize this notion is via the partial derivative matrix (sometimes called the *Jacobian*) of a mapping $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$. This is an $n \times k$ matrix denoted $\frac{\partial x}{\partial t}$ defined as follows:

$$\frac{\partial x}{\partial t} \triangleq \begin{pmatrix} \frac{\partial x_1}{\partial t_1} & \cdots & \frac{\partial x_1}{\partial t_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial t_1} & \cdots & \frac{\partial x_n}{\partial t_k} \end{pmatrix},$$

where the partial derivatives are defined in the standard way, as formal derivatives of polynomials. Let us define the *rank* of $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ to be the rank of the matrix $\frac{\partial x}{\partial t}$ when considered as a matrix over the field of rational functions in variables $t_1, \ldots, t_k$. We say that $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ is *non-degenerate* if its rank is k (since x cannot have rank larger than k).

Definition 1.1 (Polynomial Source). *Let $\mathbb{F}$ be a finite field. A distribution X over $\mathbb{F}^n$ is an* (n, k, d)-polynomial source *over $\mathbb{F}$, if there exists a non-degenerate mapping $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ such that X is sampled by choosing t uniformly at random in $\mathbb{F}^k$ and outputting $x(t)$.*

It is easy to see that the above definition of a polynomial source is indeed a generalization of the affine case, since the partial derivative matrix of an affine mapping is simply its coefficient matrix (in some basis). It is important to note that any weak source can be represented as the image of *some* polynomial mapping over a finite field $\mathbb{F}$. However, in general, the polynomials representing the source will have very high degrees (this can be seen by a simple counting argument). Since it is known [CG88] that deterministic extraction from arbitrary sources is impossible, we see that restricting our attention to low degree mappings is essential.

Rank and min-entropy: One reason for using the rank of the partial derivative matrix is that, over sufficiently large prime fields, it allows us to prove a lower-bound on the entropy of an (n, k, d)-polynomial source. This lower bound follows from a theorem of Wooley [Woo96] (see Theorem 2.8).

Roughly speaking, Wooley's theorem implies that a distribution sampled by a non-degenerate mapping $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ is close (in statistical distance) to a distribution with min-entropy at least $k \cdot \log\left(\frac{|\mathbb{F}|}{2d}\right)$. Rewriting this quantity as

$$\left(1 - \frac{\log(2d)}{\log(|\mathbb{F}|)}\right) \cdot k \cdot \log(|\mathbb{F}|),$$

we see that, as $|\mathbb{F}|$ grows, this bound 'approaches' the entropy bound of $k \cdot \log(|\mathbb{F}|)$ we have for affine sources of the same rank.

Rank and algebraic independence. Over fields of exponential characteristic (or of characteristic zero) we will see that the above notion of the rank of a mapping coincides with the more intuitive notion of *algebraic independence* (see Section 2 for the relevant definitions). Roughly speaking, over such fields, a mapping $x = (x_1, \ldots, x_n) \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ has rank k iff the set of polynomials $\{x_1(t), \ldots, x_n(t)\}$ contains k algebraically independent polynomials (we should note that the direction "rank $k \to$ algebraic independence" is true over any field, regardless of its characteristic). Since we want some of our results to hold also over fields of polynomial size we opt to use the rank of the partial derivative matrix in our definition of a polynomial source. In Section 3 we give a detailed discussion of the connection between algebraic independence and rank. Our proofs are direct extensions of the treatment appearing in [ER93] and in [L'v84] where the equivalence between the two notions is shown over the complex numbers.

1.2 Rank Extractors

The above discussion of polynomial sources raises the following natural question: Can we 'extract' the rank of these sources without destroying their structure? In other words, can we construct a *fixed* polynomial mapping $y : \mathbb{F}^n \to \mathbb{F}^k$ such that for any non-degenerate $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ the composition of y with x is a non-degenerate mapping from $\mathbb{F}^k$ to $\mathbb{F}^k$? We call a non-degenerate mapping $z : \mathbb{F}^k \to \mathbb{F}^k$ a *full rank* mapping and a mapping y satisfying the above condition a *rank extractor.*

Definition 1.2 (Rank Extractor). *Let $\mathbb{F}$ be some field. Let $y : \mathbb{F}^n \to \mathbb{F}^k$ be a polynomial mapping defined by*

$$y(x) = (y_1(x_1, \ldots, x_n), \ldots, y_k(x_1, \ldots, x_n)),$$

where each y_i is a multivariate polynomial over $\mathbb{F}$. We say that y is an (n, k, d)-rank extractor over $\mathbb{F}$ if for every non-degenerate mapping $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ the composition $y \circ x : \mathbb{F}^k \to \mathbb{F}^k$ has rank k. We will call such a mapping y explicit *if it can be computed in polynomial time.*[1]

Clearly, a construction of a rank extractor will bring us closer to constructing an extractor for low degree polynomial sources. Using an explicit rank extractor reduces the problem of constructing an extractor for arbitrary polynomial sources into the problem of constructing an extractor for polynomial sources of full rank. Surprisingly enough, the problem of extraction from full rank sources is not so easy and seems to require the use of deep results from algebraic geometry.

[1] More precisely, if it has a polynomial size arithmetic circuit that can be generated in polynomial time, given n,k and d.

Our first main result is a construction of an explicit (n, k, d)-rank extractor over $\mathbb{F}$, where $\mathbb{F}$ can be any field of characteristic zero or of characteristic at least poly(n, d). It is natural to require that the degree of the rank extractor will be as small as possible. Clearly the degree has to be larger than 1 since an affine mapping cannot be a rank extractor (because we can always 'hide' a polynomial source in the kernel of such a mapping). The rank extractors we construct have degree that is bounded by a polynomial in n and in d. In Section 4 we prove the following theorem:

Theorem 1. *Let $k \leq n$ and d be integers. Let $\mathbb{F}$ be a field of characteristic zero or of characteristic larger than $8k^2d^3n$. Then there exists an explicit (n, k, d)-rank extractor over $\mathbb{F}$ whose degree is bounded by $8k^2d^2n$. Moreover, this rank extractor can be computed in time poly$(n, \log(d))$.*

We note that our construction of rank extractors does not depend on the underlying field. We give a single construction, defined using integers, that is a rank extractor over any field satisfying the conditions of Theorem 1. We note that even if we do not restrict the degree of the rank extractor to be polynomial there does not seem to be a 'trivial' construction.

1.3 Extractors and Condensers for Polynomial Sources

As was mentioned in the previous section, applying the rank extractor given by Theorem 1 reduces the problem of constructing an extractor for (n, k, d)-polynomials sources into the problem of constructing an extractor for (k, k, d')-polynomial sources, where d' is the degree of the source obtained *after* applying the rank extractor. (Note that Theorem 1 implies that d' is polynomial in n and d). Our second main result is a construction of such an extractor. Before stating our result we give a formal definition of an extractor for polynomial sources.

Definition 1.3 (Extractor). *Let $k \leq n$ and d be integers. Let $\mathbb{F}$ be a finite field. A function $E : \mathbb{F}^n \to \{0, 1\}^m$ is a (k, d, ϵ)-*extractor for polynomial sources *if for every (n, k, d)-polynomial source X over $\mathbb{F}^n$, the random variable $E(X)$ is ϵ-close to the uniform distribution on $\{0, 1\}^m$. We say that E is* **explicit** *if it can be computed in poly$(n, \log(d))$ time.*

The following theorem, proved in Section 5, asserts the existence of an explicit extractor for full rank polynomial sources over sufficiently large prime fields. The output length of this extractor is $\Omega(k \cdot \log(|\mathbb{F}|))$, which is within a multiplicative constant of the maximal length possible. The main tool in the proof of our theorem is a result of Bombieri [Bom66] giving exponential sum estimates for polynomials defined over low degree curves.

Theorem 2. *There exists absolute constants C and c such that the following holds: Let k and $d > 1$ be integers and let $\mathbb{F}$ be a field of prime cardinality $p > d^{Ck}$. Then, there exists an explicit (k, d, ϵ)-extractor $E : \mathbb{F}^k \to \{0, 1\}^m$ for polynomial sources over $\mathbb{F}^k$ with $m = \lfloor c \cdot k \cdot \log(p) \rfloor$ and $\epsilon = p^{-\Omega(1)}$.*

Combining Theorem 2 with Theorem 1 gives an extractor for general polynomial sources. This extractor, whose existence is stated in the following corollary, also has output length which is within a multiplicative constant of optimal.

Corollary 1.4. *There exists absolute constants C and c such that the following holds: Let $k \leq n$ and $d > 1$ be integers and let $d' = 8k^2d^3n$. Let $\mathbb{F}$ be a field of prime cardinality $p > (d')^{Ck}$. Then, there exists an explicit (k, d, ϵ)-extractor $E : \mathbb{F}^n \to \{0, 1\}^m$ for polynomial sources over $\mathbb{F}^n$ with*

$m = \lfloor c \cdot k \cdot \log(p) \rfloor$ *and* $\epsilon = p^{-\Omega(1)}$.

It is possible to improve the output length of our extractors so that it is equal to a $(1-\alpha)$-fraction of the source min entropy, for any constant $\alpha > 0$. This improvement, which was suggested to us by Salil Vadhan is described in Section 6.

We note that both in Corollary 1.4 and in Theorem 2, the bound on the field size does not pose a computational problem. Over a finite field $\mathbb{F}$, arithmetic operations can be performed in time polynomial in $\log(|\mathbb{F}|)$, and hence all computations required by the extractor can be performed in polynomial time. However, it remains an interesting open problem whether extraction can be performed over smaller fields, say of size polynomial in n and in d.

Condensers Over Polynomially Large Fields: We note that over polynomially large fields, our techniques give a deterministic *condenser* for polynomial sources. A condenser is a relaxation of an extractor and is required to output a distribution with 'high' min-entropy rather than a uniform distribution. The word 'condenser' implies that the length of the output should be smaller then the length of the input. That is, the aim of a condenser is to 'compress' the source while keeping as much of the entropy as possible. For convenience, we define condensers as mappings over alphabet $\mathbb{F}$ rather than the standard definition using binary alphabet.

Definition 1.5 (Condenser). *Let $\mathcal{D}$ be a family of distributions over $\mathbb{F}^n$. A function $C : \mathbb{F}^n \rightarrow \mathbb{F}^m$ is an (ϵ, k')-*condenser *for $\mathcal{D}$ if for every X in $\mathcal{D}$ the distribution $C(X)$ is ϵ-close to having min-entropy at least k'. A condenser is* explicit *if it can be computed in polynomial time.*

From Wooley's theorem [Woo96], mentioned earlier, it follows that if we apply a rank extractor to a polynomial source then we get a source which is close to having high min-entropy. The next theorem follows immediately from Wooley's theorem (Corollary 2.9) and, in view of Theorem 1, shows the existence of explicit condensers for polynomial sources over polynomially large fields.

Theorem 3. *Let $k \leq n$ and d, d' be integers. Let $\mathbb{F}$ be a field of prime cardinality larger than $d \cdot d'$. Let $y : \mathbb{F}^n \rightarrow \mathbb{F}^k$ an (n, k, d)-rank extractor such that $\deg(y) \leq d'$. Then y is an (ϵ, k')-condenser for the family of (n, k, d)-polynomial sources over $\mathbb{F}$, where $\epsilon = \frac{d \cdot d' \cdot k}{|\mathbb{F}|}$ and $k' = k \cdot \log(|\mathbb{F}|/2dd')$.*

It should be noted that this condenser is 'almost' the best one could hope for (without building an extractor, of course). To see this, suppose that $|\mathbb{F}| \approx (2d')^c$ for some constant $c > 1$, where d' is the degree of the rank extractor. We get that the output of the condenser is close to having min-entropy

$$k' = k \cdot \log(|\mathbb{F}|/2d') \approx \left(1 - \frac{1}{c}\right) \cdot k \cdot \log(|\mathbb{F}|),$$

and so the ratio between the length of the output (in bits) and its min-entropy can be made arbitrarily close to one by choosing c to be large enough.

Dispersers Over the Complex Field. A disperser is a relaxation of an extractor in which the output is only required to have large support (instead of being close to uniform). Dispersers are usually considered only for distributions over finite sets. However, for polynomial sources we can extend our view also for infinite sets (namely infinite fields). It is shown in [ER93] that the image of a full rank mapping $x \in \mathcal{M}(\mathbb{C}^k \to \mathbb{C}^k, d)$ contains all of $\mathbb{C}^k$ except for the zero set of some polynomial. This shows that our rank extractors can be viewed as deterministic *dispersers* for polynomial sources over $\mathbb{C}$. That is, a rank extractor is a fixed polynomial transformation mapping *any* polynomial source into almost all of $\mathbb{C}^k$. We discuss this observation in Section 8.

1.4 Rank Versus Entropy - Weak Polynomial Sources

So far we focused on extraction from sources which were defined algebraically - we were given a bound on the algebraic rank of the set of polynomials we extract from. We now switch to the more standard definition (from the extractor literature standpoint) of extraction from sources with given min-entropy (see Definition 2.2). These will be called *Weak Polynomial Sources.*

Definition 1.6 (Weak Polynomial Source). *A distribution X over $\mathbb{F}^n$ is an (n, k, d)-weak polynomial source (WPS) if*

- *There exists a polynomial mapping $x \in \mathcal{M}(\mathbb{F}^n \to \mathbb{F}^n, d)$ (of arbitrary rank) such that X is sampled by choosing t uniformly in $\mathbb{F}^n$ and outputting $x(t)$.*
- *X has min entropy at least $k \cdot \log(|\mathbb{F}|)$.*

Notice in the definition that the min-entropy threshold is $k \cdot \log(|\mathbb{F}|)$ (instead of just k). This is to hint to the connection (which we prove later) between the rank of the source and its entropy. Intuitively, a distribution sampled by a rank r mapping $x : \mathbb{F}^n \to \mathbb{F}^n$ "should" have entropy roughly $r \cdot \log(|\mathbb{F}|)$ and indeed, for affine sources, this is exactly the case.

The following theorem, whose proof can be found in Section 7, shows the existence of an explicit deterministic extractor for the class of weak polynomial sources (an extractor for weak polynomial sources is defined in an analogous fashion to Definition 1.3)

Theorem 4. *There exists absolute constants C and c such that the following holds: Let $k \leq n$ and $d > 1$ be integers and let $d' = 8k^2d^3n$. Let $\mathbb{F}$ be a field of prime cardinality $p > (d')^{Ck}$. Then, there exists an explicit (k, d, ϵ)-extractor $E : \mathbb{F}^n \to \{0,1\}^m$ for weak polynomial sources over $\mathbb{F}^n$ with $m = \lfloor c \cdot k \cdot \log(p) \rfloor$ and $\epsilon = p^{-\Omega(1)}$.*

The parameters of the extractor given by the theorem can be seen to be roughly the same as those of the extractor for regular polynomial sources (Corollary 1.4). In fact, the extractor we use for weak polynomial sources is the same one we used for polynomial sources. The proof of Theorem 4 will follow by showing that any (n, k, d)-WPS is close (in statistical distance) to a convex combination of (n, k, d)-polynomial sources. This implies that any extractor that works for polynomial sources will work also for weak polynomial sources.

The Entropy of a Polynomial Mapping: We can use the methods employed in the proof of Theorem 4 to show that over sufficiently large fields, the output of a low degree polynomial mapping $x \in \mathcal{M}(\mathbb{F}^n \to \mathbb{F}^n, d)$ is always close to having entropy approximately $\mathrm{rank}(x) \cdot \log(|\mathbb{F}|)$. This can be viewed as a generalization of the simple fact that for an *affine* mapping x, the entropy is always equal to $\mathrm{rank}(x) \cdot \log(|\mathbb{F}|)$. (See Section 7.2 for the formal statement of this result.)

Extractors for Poly-Size Arithmetic Circuits: An interesting corollary of Theorem 4 is the existence of deterministic extractors for the class of distributions sampled by polynomial sized arithmetic circuits over exponentially large fields. This follows from the fact that the degrees of the polynomials computed by poly-size circuits are exponential, and the construction of an (n, k, d)-rank extractor is efficient even when d is exponential.

We say that a distribution X on $\mathbb{F}^n$ is sampled by a size s arithmetic circuit if there exists an arithmetic circuit A of size s with n inputs and n outputs such that the fan-in of each gate is at most two and such that X is the distribution of the output of A on a random input, chosen uniformly from $\mathbb{F}^n$. We say that X is an (n, k, s)-*arithmetic source* if X is sampled by a size s arithmetic circuit and its min-entropy is at least $k \cdot \log(|\mathbb{F}|)$.

Corollary 1.7. *There exists absolute constants C and c such that the following holds: Let $k \leq n$ and $s > 1$ be integers. Let $d = 2^s$ and let $d' = 8k^2d^3n$. Let $\mathbb{F}$ be a field of prime cardinality $p > (d')^{Ck}$. Then, there exists an explicit function $E : \mathbb{F}^n \to \{0,1\}^m$ such that for every (n, k, s)-arithmetic source X over $\mathbb{F}$, the distribution of $E(X)$ is ϵ-close to uniform, where $m = \lfloor c \cdot k \cdot \log(p) \rfloor$ and $\epsilon = p^{-\Omega(1)}$. That is, E is an extractor for the class of (n, k, s)-arithmetic sources.*

It is interesting to contrast this result to the extractors of [TV00] from polynomial size *boolean* circuits. Their extractors rely on complexity assumptions, and they prove that such assumptions are necessary. It is interesting that over large fields no such assumptions, nor lower bounds, are necessary.

1.5 Organization

Section 2 contains general preliminaries on probability distributions and finite field algebra. Section 3 contains a detailed discussion on the connection between algebraic independence and rank. In Section 4 we describe our construction of a rank extractor and prove Theorem 1. In Section 5 we construct and analyze an extractor for full rank polynomial sources and prove Theorem 2. In Section 6 we show how to increase the output length of our extractors. In Section 7 we discuss extractors for weak polynomial sources and prove Theorem 4. In Section 8 we discuss rank extractors over the complex numbers. Appendix A contains background from Algebraic Geometry required for the proof of Theorem 2.

2 General Preliminaries

2.1 Probability Distributions

Let Ω be some finite set. Let P be a distribution on Ω. For $B \subseteq \Omega$, we denote the probability of B according to P, by $\Pr_P(B)$ or $\Pr(P \subseteq B)$; When $B \in \Omega$, we will also use the notation $\Pr(P = B)$.

Given a function $A : \Omega \to U$, we denote by $A(P)$ the distribution induced on U when sampling t by P and calculating $A(t)$. When we write $t_1, \ldots, t_k \leftarrow P$, we mean that $t_1, \ldots, t_k$ are chosen *independently* according to P. We denote by U_Ω the uniform distribution on Ω. Given a function $x : \mathbb{F}^m \to \mathbb{F}$, we denote by $x(U_m)$ the distribution $x(U_{\mathbb{F}^m})$. For a distribution P on Ω^d and $j \in [d]$, we denote by P_j the marginal distribution of P on the j'th coordinate.

The *statistical distance* between two distributions P and Q on Ω, denoted by $|P - Q|$, is defined as

$$|P - Q| \triangleq \max_{S \subseteq \Omega} \left| \Pr_P(S) - \Pr_Q(S) \right| = \frac{1}{2} \sum_{w \in \Omega} \left| \Pr_P(w) - \Pr_Q(w) \right|.$$

We say that P is *ϵ-close* to Q, denoted $P \overset{\epsilon}{\sim} Q$, if $|P - Q| \leq \epsilon$. We denote the fact that P and Q are identically distributed by $P \sim Q$. The following Lemma is trivial:

Lemma 2.1. *Let P, V be distributions on a set Ω. Suppose, $P = \delta \cdot R + (1-\delta) \cdot V$, for two distributions R and V and $0 < \delta < 1$. Then $P \overset{\delta}{\sim} V$.*

We use *min-entropy* to measure the amount of randomness in a given distribution:

Definition 2.2 (Min-entropy). *Let X be a distribution over a finite set Γ. The min-entropy of X is defined as*

$$H_\infty(X) \triangleq \min_{x \in supp(X)} \log \left(\frac{1}{\mathbf{Pr}[X = x]} \right).$$

Another useful measure of entropy is *collision probability.*

Definition 2.3 (Collision Probability). *Let X be a distribution over a finite set Γ. The* collision probability *of X is defined as*

$$cp(X) \triangleq \sum_{x \in supp(X)} \mathbf{Pr}[X = x]^2 = \mathbf{Pr}_{x_1, x_2 \leftarrow X}[x_1 = x_2]$$

The following lemma gives us a quantitative translation between the two quantities of min entropy and collision probability.

Lemma 2.4 (Lemma 3.6 in [BIW04]). *Let X be a distribution over a finite set Γ. Suppose that $cp(X) \leq \frac{1}{a \cdot b}$. Then X is $\frac{1}{\sqrt{a}}$-close to a distribution with min entropy at least $\log(b)$.*

2.2 Polynomials Over Finite Fields

We review some basic notions regarding polynomials defined over finite fields. Readers not familiar with the subject can find a more comprehensive treatment in [LN97]. For a field $\mathbb{F}$ we denote by $\mathbb{F}[t_1, \ldots, t_k]$ the ring of polynomials in k-variables $t_1, \ldots, t_k$ with coefficients in $\mathbb{F}$. We denote by $\mathbb{F}(t_1, \ldots, t_k)$ the field of rational functions in variables $t_1, \ldots, t_k$. We denote by $\deg(f)$ the total degree of f and by $\deg_{t_j}(f)$ the degree of f as a polynomial in t_j. We write $f \equiv 0$ or $f(t) \equiv 0$ if f is the zero polynomial (all coefficients of f are zero). Note that over the finite field $\mathbb{F}$ of prime cardinality p, the polynomial $f(t) = t^p - t$ is **not** the zero polynomial, even though $f(a) = 0$ for all $a \in \mathbb{F}$.

We say that the polynomials $f_1, \ldots, f_m \in \mathbb{F}[t_1, \ldots, t_k]$ are *algebraically dependent* if there exists a non-zero polynomial $h \in \mathbb{F}[z_1, \ldots, z_m]$ such that $h(f_1(t), \ldots, f_m(t)) \equiv 0$. We sometimes refer to this polynomial h as the *annihilating polynomial* of $f_1, \ldots, f_m$. We say that $f_1, \ldots, f_m$ are *algebraically independent* if such a polynomial h does not exist.

For a polynomial $f \in \mathbb{F}[t_1, \ldots, t_k]$ we denote by $\frac{\partial f}{\partial t_j} \in \mathbb{F}[t_1, \ldots, t_k]$ the formal partial derivative of f with respect to the variable t_j. When using derivatives over a finite field we should be careful of 'strange' behavior of the derivative. For example, the derivative of t^p over a field of characteristic p is equal to zero. This is 'strange' since t^p is not a constant function (in fact, it is a permutation). The following claim, which we use implicitly in many of our proofs, describes the exact conditions under which this 'strange' behavior happens.

Claim 2.5. *Let $\mathbb{F}$ be a field of characteristic p and let $f \in \mathbb{F}[t_1, \ldots, t_k]$ and $j \in [k]$ be such that $\frac{\partial f}{\partial t_j} \equiv 0$. Then all degrees of t_j appearing in f are multiples of p. In particular, if $\deg_{t_j}(f) < p$. Then $\frac{\partial f}{\partial t_j} \equiv 0$ iff $\deg_{t_j}(f) = 0$.*

For a vector of polynomials $\bar{f} = (f_1, \ldots, f_m) \in (\mathbb{F}[t_1, \ldots, t_k])^m$ we can define the *partial derivative matrix* of $\bar{f}$ as

$$\frac{\partial \bar{f}}{\partial t} \triangleq \begin{pmatrix} \frac{\partial f_1}{\partial t_1} & \cdots & \frac{\partial f_1}{\partial t_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial t_1} & \cdots & \frac{\partial f_m}{\partial t_k} \end{pmatrix}.$$

We denote by $\mathrm{rank}(\bar{f})$ the rank, over $\mathbb{F}(t_1, \ldots, t_k)$, of the matrix $\frac{\partial \bar{f}}{\partial t}$.

Another useful property of polynomials, which we will use often, is the bound on the number of roots they can have. This generalization of the fundamental theorem of algebra is due to Schwartz and Zippel [Sch80, Zip79].

Lemma 2.6 (Schwartz-Zippel). *Let $\mathbb{F}$ be a field and let $f \in \mathbb{F}[t_1, \ldots, t_k]$ be a non zero polynomial with $\deg(f) \leq d$. Then, for any finite subset $S \subset \mathbb{F}$ we have*

$$\left|\left\{c \in S^k \,:\, f(c) = 0\right\}\right| \leq d \cdot |S|^{k-1}.$$

A simple corollary of the Schartz-Zippel Lemma is the following Claim:

Claim 2.7. *Let $\mathbb{F}$ be a finite field and let $f \in \mathbb{F}[t_1, \ldots, t_k]$ be a polynomial of total degree at most d. Fix any $1 < i \leq k$. For $c = (c_i, \ldots, c_k) \in \mathbb{F}^{k-i+1}$ define*

$$f_c(t_1, \ldots, t_{i-1}) \triangleq f(t_1, \ldots, t_{i-1}, c_i, \ldots, c_k)$$

Then

$$\Pr_{c \leftarrow \mathbb{F}^{k-i+1}}(f_c \equiv 0) \leq \frac{d}{|\mathbb{F}|}$$

2.3 The Number of Solutions to a System of Polynomial Equations

We will use a version of Bezout's Theorem proved by Wooley [Woo96]. This theorem, mentioned informally in the introduction, will give us a connection between algebraic rank and min entropy. We note that the formulation of Wooley's theorem stated here is weaker then the original formulation appearing in [Woo96] (the original form of the theorem speaks of congruences modulo p^s for any s).

Theorem 2.8 (Rephrased from Theorem 1 in [Woo96]). *Let $\mathbb{F}$ be a field of prime cardinality p. Let k and d be integers. Let $x = (x_1, \ldots, x_k) \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^k, d)$ be such that rank$(x) = k$ and denote by $J(t) \triangleq \det\left(\frac{\partial x}{\partial t}\right)(t)$. For $a \in \mathbb{F}^k$ let*

$$N_a \triangleq \left|\left\{c \in \mathbb{F}^k \quad : \quad x(c) = a \quad and \quad J(c) \neq 0\right\}\right|.$$

Then for every $a \in \mathbb{F}^k$, $N_a \le d^k$.

We can interpret this theorem as saying that a distribution X sampled by a non-degenerate mapping $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^k, d)$ is close to a distribution with high min-entropy, where the closeness is related to the number of zeros of the determinant of $\frac{\partial x}{\partial t}$. Since this determinant is a non-zero low-degree polynomial, we get that the distance from the high min-entropy distribution is small. This is stated more precisely by the following Corollary, which also extends our view to mappings in $\mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ for $k \le n$.

Corollary 2.9. *Let $\mathbb{F}$ be a field of prime cardinality. Let $k \le n$ and d be integers such that $|\mathbb{F}| > 2dk$. Let X be an (n, k, d)-polynomial source over $\mathbb{F}$. Then X is ϵ-close to a distribution with min-entropy at least $k \cdot \log\left(\frac{|\mathbb{F}|}{2d}\right)$, where $\epsilon = \frac{d \cdot k}{|\mathbb{F}|}$.*

Proof. X is the distribution $x(U_k)$ for a non-degenerate mapping $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$. Since x has rank k the matrix $\frac{\partial x}{\partial t}$ has a non-singular square sub-matrix. W.l.o.g assume that this matrix is composed of the first k rows of $\frac{\partial x}{\partial t}$. Let us also denote the determinant of this sub-matrix as $J(t)$.

Denote by C the event that $J(t) = 0$ and let $\delta = \Pr_{t \leftarrow \mathbb{F}^k}(C)$. Write X as a convex combination of conditional distributions as follows

$$X = \delta \cdot (X|C) + (1 - \delta) \cdot (X|\neg C).$$

Note that, since $J(t)$ is a non-zero polynomial of degree at most $d \cdot k$, by Lemma 2.6 we have that $\delta \le \frac{d \cdot k}{|\mathbb{F}|}$.

We claim that the distribution $(X|\neg C)$ has min-entropy at least $k \cdot \log(|\mathbb{F}|/2d)$: For any $a \in \mathbb{F}^n$, using Theorem 2.8

$$\Pr(X = a|\neg C) = \frac{\Pr(X = a \wedge \neg C)}{1 - \delta} \le \frac{d^k}{|\mathbb{F}|^k \cdot (1 - \delta)}$$

$$\le \frac{d^k}{|\mathbb{F}|^k \cdot (1 - dk/|\mathbb{F}|)} \le \frac{2d^k}{|\mathbb{F}|^k} \le \left(\frac{2d}{|\mathbb{F}|}\right)^k,$$

(here we use the bound on $|\mathbb{F}|$). Thus, $(X|\neg C)$ has min-entropy at least $k \cdot \log(|\mathbb{F}|/2d)$ and using Lemma 2.1 we are done. □

3 Algebraic Independence and Rank

In [ER93] it is shown that, over the complex numbers, the two notions of rank and algebraic independence are equivalent. That is, the polynomials $x_1, \ldots, x_r \in \mathbb{F}[t_1, \ldots, t_k]$ are algebraically independent iff the matrix $\frac{\partial x}{\partial t}$ has maximal rank. In this section we prove two theorems showing that this connection is also valid over finite fields, provided the characteristic of the field is sufficiently large. We start by showing that maximal rank implies algebraic independence. This direction does not require the field characteristic to be large.

Theorem 3.1. *Let $\mathbb{F}$ be a field of characteristic p. Let $x = (x_1, \ldots, x_r) \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^r, d)$ for some d, where $r \le k$. If x has rank r then $x_1, \ldots, x_r$ are algebraically independent.*

Proof. Assume for contradiction, that $x_1, \ldots, x_r$ are algebraically dependent. Let $g(z_1, \ldots, z_r)$ be a non zero polynomial of minimal degree such that $g(x_1(t), \ldots, x_r(t)) \equiv 0$. Denote $g_i = \frac{\partial g}{\partial z_i}$.

Claim 3.2. *For some $1 \leq i \leq k$, g_i is non-zero.*

Proof. Fix some $1 \leq i \leq k$. Assume that $g_i \equiv 0$. Then, by Claim 2.5, all non-zero powers of z_i in g are multiples of p. Assume for contradiction that for all i, $g_i \equiv 0$. Then $g = h^p$ for some $h(z_1, \ldots, z_r)$, and

$$(h(x_1(t), \ldots, x_r(t)))^p \equiv 0 \Rightarrow h(x_1(t), \ldots, x_r(t)) \equiv 0,$$

and this is a contradiction to the minimality of g. □

We will go on to show that the derivatives of g form a non trivial vector which is orthogonal to all the columns of $\frac{\partial x}{\partial t}$, contradicting our assumption that $\frac{\partial x}{\partial t}$ has maximal rank. Using the above claim, fix an i such that g_i is non-zero. By the minimality of the degree of g we know that $g_i(x_1(t), \ldots, x_r(t))$ is non-zero as a polynomial in t (the degree of the derivative is always smaller than that of the original polynomial). Define $\overline{g}(t) \triangleq g(x_1(t), \ldots x_r(t))$. Note that $\overline{g}(t) \equiv 0$. Using the chain rule, for $1 \leq j \leq k$ we have

$$0 = \frac{\partial \overline{g}}{\partial t_j} = \sum_{l=1}^{r} g_l(x(t)) \cdot \frac{\partial x_l}{\partial t_j}.$$

Note that the rightmost expression is the inner product of the non-zero vector

$$u = (g_1(x(t)), \ldots, g_r(x(t)))$$

and the j'th column of the matrix $\frac{\partial x}{\partial t}$. Thus, we have

$$u \cdot \frac{\partial x}{\partial t} = 0$$

for $u \neq 0$ and so the rank of $\frac{\partial x}{\partial t}$ is at most $r - 1$, a contradiction. □

We now turn to prove the other direction, which states that algebraic independence implies maximal rank. In order to prove this direction we require the field characteristic to be larger than $(k+1)d^k$ where k is the number of variables and d is the total degree of the polynomials. This requirement stems from the degree of the annihilating polynomial we find in the proof. Our proof is based on the same ideas appearing in [ER93, L'v84, Woo96]. We are not aware how tight is the degree bound we get in the proof. Another approach is to use Grobner Bases, which often leads to double exponential degrees.

Theorem 3.3. *Let $\mathbb{F}$ be a field of characteristic p. Let d, k and n be integers such that $p > D$, where $D = (k+1) \cdot d^k$. Let $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ have rank smaller than n. Then, there exists a non zero polynomial $h \in \mathbb{F}[z_1, \ldots, z_n]$ of total degree at most D such that*

$$h(x_1(t), \ldots, x_n(t)) \equiv 0.$$

Proof. Fix any d and k. We first prove the theorem for $n \geq k+1$. Assume w.l.g. that $n = k+1$ (if $n > k+1$ we can use this case to find an h that uses only the first $k+1$ variables). In this case, the coefficients of the required h can be found by showing that a certain system of linear equations has more degrees of freedom than constraints. More precisely, we want a non-zero polynomial h of degree at most D such that $\overline{h}(t) \triangleq h(x_1(t), \ldots, x_n(t)) \equiv 0$. The number of constraints is the number of coefficients of $\overline{h}$. Since $deg(\overline{h}) \leq d \cdot D$, this is at most $\binom{d \cdot D+k}{k}$. The number of variables is the number of coefficients of h which is $\binom{D+n}{n} = \binom{D+k+1}{k+1}$. We show that the number of variables is larger than the number of constraints:

$$\binom{D+k+1}{k+1} \Big/ \binom{d \cdot D+k}{k} = \frac{(D+k+1)!}{D!(k+1)!} \cdot \frac{k!(d \cdot D)!}{(d \cdot D+k)!}$$
$$= \frac{(D+1)\cdots(D+k+1)}{(k+1)\cdot(d\cdot D+1)\cdots(d \cdot D+k)} \geq \left(\frac{D}{d \cdot D}\right)^k \cdot \frac{D+k+1}{k+1}$$
$$= \frac{D+k+1}{d^k \cdot (k+1)} > 1.$$

We now prove the claim for $n \leq k$ by backwards induction on n. We assume the claim for $n+1$ and prove it for n. Assume for contradiction, that there is no non-zero polynomial $h(z_1, \ldots, z_n)$ of degree at most D such that $h(x_1(t), \ldots, x_n(t)) \equiv 0$. Using the induction hypothesis, for each $1 \leq i \leq k$ we have a non-zero polynomial $h_i(z_1, \ldots, z_n, w)$ of degree at most D with

$$h_i(x_1(t), \ldots, x_n(t), t_i) \equiv 0. \tag{1}$$

We will go on to show that the partial derivatives of the polynomials h_i form a matrix which is the 'inverse' of $\frac{\partial x}{\partial t}$, contradicting our assumption about the rank of $\frac{\partial x}{\partial t}$. W.l.o.g assume that h_i is a minimal degree polynomial satisfying (1). For $1 \leq j \leq n$ denote $h_{i,j} = \frac{\partial h_i}{\partial z_j}$ and denote $h_{i,0} = \frac{\partial h_i}{\partial w}$. By our contradiction assumption, h_i must contain non-zero powers of w, and since $deg(h_i) < p$ this implies that $h_{i,0}$ is non-zero. By the minimality of the degree of h_i, we have that $h_{i,0}(x_1(t), \ldots, x_n(t), t_i)$ is a *non-zero* polynomial in t. Taking the derivative of (1) for each $1 \leq l \leq k$, we have

$$0 = \sum_{j=1}^{n} h_{i,j} \cdot \frac{\partial x_j}{\partial t_l} + \delta_{i,l} \cdot h_{i,0}.$$

Since we can divide by the non-zero $h_{i,0}$ we get

$$\frac{-1}{h_{i,0}} \sum_{j=1}^{n} h_{i,j} \cdot \frac{\partial x_j}{\partial t_l} = \delta_{i,l}$$

for every $1 \leq i \leq k$ and $1 \leq l \leq k$. Therefore, we have $H \cdot \frac{\partial x}{\partial t} = I$, where H is the $k \times n$ matrix with $H_{i,j} = \frac{-h_{i,j}}{h_{i,0}}$, contradicting the assumption that $\frac{\partial x}{\partial t}$ has rank smaller than n. □

4 An Explicit Rank Extractor

In this section we describe our construction of a rank extractor and prove Theorem 1.

Construction 1. *Let $k \leq n$ and d be integers. Let $s_2 = dk+1$ and $s_1 = (2dn+1) \cdot s_2$. Let $l_{ij} = i \cdot (s_1 + j \cdot s_2)$. Define for each $1 \leq i \leq k$*

$$y_i(x) = y_i(x_1, \ldots, x_n) \triangleq \sum_{j=1}^{n} \frac{1}{l_{ij}+1} \cdot x_j^{l_{ij}+1}.$$

Let $y = (y_1, \ldots, y_k)$ be the output of the construction. Notice that $y(x)$ is defined in such a way that the partial derivative $\frac{\partial y_i}{\partial x_j}$ is exactly $x_j^{l_{ij}}$.

We prove the following theorem, which directly implies Theorem 1.

Theorem 4.1. *Let $\mathbb{F}$ be a field of characteristic zero or of characteristic larger than $d' = 8k^2d^3n$. Let $x \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ be of rank k. Let $y : \mathbb{F}^n \to \mathbb{F}^k$ be as in Construction 1. Then the composition $(y \circ x)(t)$ is in $\mathcal{M}(\mathbb{F}^k \to \mathbb{F}^k, d')$ and has rank k.*

4.1 Preliminaries For The Proof Of Theorem 4.1

4.1.1 Sums of Powers of Polynomials

The following lemma shows how to pick integers $c_1, \ldots, c_n$ in such a way that for any set of n polynomials $x_1(t), \ldots, x_n(t)$ of bounded degree, the polynomials $x_1(t)^{c_1}, \ldots, x_n(t)^{c_n}$ will have degrees that are different by at least some fixed number.

Lemma 4.2. *Let $x_1(t), \ldots, x_n(t)$ be k-variate non-constant polynomials over some field $\mathbb{F}$. Denote by $d_i > 0$ the degree of the polynomial x_i. Let $d \geq \max_i\{d_i\}$. Let A and B be two positive integers such that $A \geq (2dn+1) \cdot B$ and let $c_i \triangleq A + Bi$ for $i \in [n]$. Then, for every $1 \leq i < j \leq n$, we have*

$$|\deg(x_i(t)^{c_i}) - \deg(x_j(t)^{c_j})| = |d_i \cdot c_i - d_j \cdot c_j| \geq B.$$

Proof. Let $1 \leq i < j \leq n$. First, suppose that $d_i = d_j$. In this case we have

$$d_j \cdot c_j - d_i \cdot c_i = d_j(A + Bj) - d_i(A + Bi) = d_j \cdot B \cdot (j - i) \geq B.$$

Next suppose $d_j \neq d_i$. In this case we have

$$\begin{aligned} |d_j \cdot c_j - d_i \cdot c_i| &= |d_j(A + Bj) - d_i(A + Bi)| \\ &= |(d_j - d_i)A + d_jBj - d_iBi| \\ &\geq |d_j - d_i|A - |d_jBj| - |d_iBi| \\ &\geq A - 2dnB \geq B. \end{aligned}$$

□

4.1.2 The Cauchy-Binet Formula

The Cauchy-Binet formula gives the determinant of the product of a $k \times n$ matrix with an $n \times k$ matrix (for $k \leq n$). Let $k \leq n$. Let A be a $k \times n$ matrix and B an $n \times k$ matrix. For a set $I \subset [n]$ of size k we denote by A_I the $k \times k$ sub-matrix of A composed of the columns of A whose indices appear in I. Similarly, we denote by B_I the sub-matrix of B composed of the rows of B whose indices are in I. The proof of the following formula can be found in [Gan59].

Lemma 4.3 (Cauchy-Binet). *Let $k \leq n$. Let A be a $k \times n$ matrix and B an $n \times k$ matrix over a field $\mathbb{F}$. Using the above notations we have*

$$\det(A \cdot B) = \sum_{\substack{I \subset [n] \\ |I| = k}} \det(A_I) \cdot \det(B_I).$$

4.2 Proof of Theorem 4.1

Let $k \leq n$, d be integers. Let $\mathbb{F}$ be a field of characteristic zero or of characteristic larger than $d' = 8k^2d^3n$. Let $x = (x_1, \ldots, x_n) \in \mathcal{M}(\mathbb{F}^k \to \mathbb{F}^n, d)$ be such that $\text{rank}(x) = k$. Let $y : \mathbb{F}^n \to \mathbb{F}^k$ be defined as in Construction 1, that is

$$y_i(x) = y_i(x_1, \ldots, x_n) \triangleq \sum_{j=1}^{n} \frac{1}{l_{ij}+1} \cdot x_j^{l_{ij}+1}, \tag{2}$$

where

$$l_{ij} = i \cdot (s_1 + j \cdot s_2)$$

$$s_1 = (2dn+1) \cdot s_2 \quad , \quad s_2 = dk+1$$

It is easy to verify that the degree of the mapping y is bounded by $8k^2d^2n$. Therefore, the degree of the composition $(y \circ x)(t)$ is bounded by $d' = 8k^2d^3n$. Therefore, since the characteristic of $\mathbb{F}$ is larger than d' (or is zero), for the rest of the proof we don't need to worry about non constant polynomials becoming zero after we take their derivative (see Claim 2.5).

Our goal is to show that the composition $y \circ x$ has rank k. In order to prove this we need to show that the determinant of the partial derivatives matrix of the composition is non zero. Write $y(t)$ to denote $y(x(t))$ and let $\frac{\partial y}{\partial t}$ denote the $k \times k$ partial derivative matrix of the mapping $y(t)$. Using the chain rule we have that

$$\frac{\partial y}{\partial t} = \frac{\partial y}{\partial x} \cdot \frac{\partial x}{\partial t},$$

where $\frac{\partial y}{\partial x}$ is a $k \times n$ matrix and $\frac{\partial x}{\partial t}$ is an $n \times k$ matrix. All the elements in these two matrices are polynomials in t, since we evaluate $\frac{\partial y}{\partial x}$ at $x = x(t)$.

Consider the element at position (i,j) in the matrix $\frac{\partial y}{\partial x}$. Taking the derivative of (2) with respect to x_j we get that

$$\frac{\partial y_i}{\partial x_j} = x_j(t)^{l_{ij}} = x_j(t)^{i \cdot (s_1 + js_2)}.$$

The Vandermonde structure of $\frac{\partial y}{\partial x}$ becomes more apparent by denoting

$$r_j(t) \triangleq x_j(t)^{s_1 + js_2}.$$

We now have that the (i,j)'th element of $\frac{\partial y}{\partial x}$ is $r_j(t)^i$. That is

$$\frac{\partial y}{\partial x} = \begin{pmatrix} r_1(t) & r_2(t) & \cdots & \cdots & r_n(t) \\ r_1(t)^2 & r_2(t)^2 & \ddots & & r_n(t)^2 \\ \vdots & \vdots & & \ddots & \vdots \\ r_1(t)^k & r_2(t)^k & \cdots & \cdots & r_n(t)^k \end{pmatrix}.$$

To facilitate writing, let us denote by $R \triangleq \frac{\partial y}{\partial x}$ and $D \triangleq \frac{\partial x}{\partial t}$. We can also assume w.l.o.g that

$$\deg(r_1(t)) \leq \ldots \leq \deg(r_n(t)), \tag{3}$$

(we let $\deg(0) = 0$) since applying the same permutation on the rows of R and on the columns of D will not change the determinant of $R \cdot D$. Now, from Lemma 4.3 (Cauchy-Binet) and using the notations of Section 4.1.2 we have that

$$\det\left(\frac{\partial y}{\partial t}\right) = \det(R \cdot D) = \sum_{\substack{I \subset [n] \\ |I| = k}} \det(R_I) \cdot \det(D_I) \tag{4}$$

Notice that if $r_i(t)$ is constant, then $x_i(t)$ is also constant and so the i'th row of the matrix D is zero. Therefore, $\det(D_I) = 0$ for every I that contains an index i such that $r_i(t)$ is constant. In view of (4) and this last observation, we can assume w.l.o.g that for all $i \in [n]$, $r_i(t)$ is non constant. (Notice that since D has maximal rank, we have at least k indices in $[n]$ for which $x_i(t)$ is non constant and so the condition $n \geq k$ is maintained).

The next three claims will show that there exist a unique set I in the above sum for which the degree of $\det(R_I) \cdot \det(D_I)$ is maximal. This will conclude the proof, since then we will have that $\det\left(\frac{\partial y}{\partial t}\right)$ is non zero, as required.

We start with a simple claim showing that the degrees of the polynomials $r_i(t)$ have large gaps between them.

Claim 4.4. *Let $r_1(t), \ldots, r_n(t)$ be the polynomials defined above. Then for every $i \in [n-1]$ we have*

$$\deg(r_{i+1}(t)) > \deg(r_i(t)) + dk.$$

Proof. Recall that $r_i(t) = x_i(t)^{s_1 + j \cdot s_2}$ and that $s_1 \geq (2dn+1) \cdot s_2$. Using Lemma 4.2 we get that

$$|\deg(r_{i+1}(t)) - \deg(r_i(t))| \geq s_2 > dk.$$

Using (3) the claim follows. □

Let $I \subset [n]$ be such that $|I| = k$. We denote by

$$d_I \triangleq \deg\left(\det(R_I)\right).$$

The next claim gives a convenient formula for d_I.

Claim 4.5. *Let $I \subset [n]$, $I = \{i_1 < \ldots < i_k\}$. Then*

$$d_I = \sum_{j=1}^{k} j \cdot \deg\left(r_{i_j}(t)\right).$$

Proof. Using the Vandermonde structure of the matrix R_I we get that

$$\det(R_I) = \prod_{j=1}^{k} r_{i_j}(t) \prod_{1 \leq j_1 < j_2 \leq k} \left(r_{i_{j_1}}(t) - r_{i_{j_2}}(t)\right).$$

In view of (3), the degree of the highest monomial in $\det(R_I)$ is obtained by multiplying k copies of $r_{i_k}(t)$ with $k-1$ copies of $r_{i_{k-1}}(t)$ and so on. This will give a monomial with degree $\sum_{j=1}^{k} j \cdot \deg(r_j(t))$. □

Define

$$\Gamma \triangleq \{ I \subset [n] \mid |I| = k \, , \, \det(D_I) \neq 0 \} \, .$$

The next and final claim shows that there exists a **unique** $I \in \Gamma$ with maximal d_I. The proof uses standard techniques from matroid theory.

Claim 4.6. *Let* $d_{\max} \triangleq \max_{I \in \Gamma}\{d_I\}$*. Then there exists a unique* $I^* \in \Gamma$ *such that* $d_{I^*} = d_{\max}$*. Moreover, for every* $I \neq I^*$ *we have that* $d_I < d_{I^*} - dk$*.*

Proof. Let $v_1, \ldots, v_n$ denote the rows of D. We can treat $v_1, \ldots, v_n$ as vectors in a k-dimensional vector space over the field of rational functions in variables $t_1, \ldots, t_k$.

We are going to construct the set I^* using the following greedy algorithm: Start with $I^* = \emptyset$ and at each step add to I^* the largest $i \in [n]$ for which the set $\{ \, v_j \mid j \in I^* \cup \{i\} \, \}$ is linearly independent. Since we assumed that D has maximal rank, this process will end after precisely k steps, yielding a set I^* of size k and such that $\det(D_{I^*}) \neq 0$. Denote by $I^* = \{i_1^* < \ldots < i_k^*\}$.

Observing the formula for d_I given by Claim 4.5 and recalling from Eq. 3 that the degrees of the polynomials r_i are strictly increasing, we see that the greedy construction of I^* ensures that $d_{I^*} = d_{\max}$. Assume for contradiction that there exists a set $I' \neq I$ in Γ such that $d_{I'} = d_{\max}$ and denote by $I' = \{i'_1 < \ldots < i'_k\}$. From the monotonicity of $\deg(r_i(t))$ it follows that there must be an index $j \in [k]$ such that $i'_j > i^*_j$ (otherwise we would have $d_{I'} < d_{I^*}$). Let $j' \in [k]$ be the largest index such that $i'_{j'} > i^*_{j'}$. Since $I' \in \Gamma$ we have that the set $\left\{ v_{i'_{j'}}, v_{i'_{j'+1}}, \ldots, v_{i'_k} \right\}$ is linearly independent. Therefore there must be an index $0 \leq \alpha \leq k - j'$ such that the vector $v_{i'_{j'+\alpha}}$ is not spanned by the set of vectors $\left\{ v_{i^*_{j'+1}}, v_{i^*_{j'+2}}, \ldots, v_{i^*_k} \right\}$. This contradicts the greedy construction of I^* since, by construction, all the vectors $v_{i^*_{j'}+1}, v_{i^*_{j'}+2}, \ldots, v_n$ are spanned by $\left\{ v_{i^*_{j'+1}}, v_{i^*_{j'+2}}, \ldots, v_{i^*_k} \right\}$.

To prove the 'moreover' part of the claim we use Claim 4.4. Let $I = \{i_1 < \ldots < i_k\}$ be such that $I \neq I^*$ and $I \in \Gamma$. Using the same logic as above we can deduce that for all $j \in [k]$, $i_j \leq i^*_j$ and that for some $j' \in [k]$, $i_{j'} < i^*_{j'}$. Plugging this information into the formula for d_I we get that

$$\begin{aligned} d_{I^*} - d_{I'} &= \sum_{j=1}^{k} j \cdot \left(\deg\left(r_{i^*_j}(t) \right) - \deg\left(r_{i_j}(t) \right) \right) \\ &\geq \deg\left(r_{i^*_{j'}}(t) \right) - \deg\left(r_{i_{j'}}(t) \right) \\ &> dk, \end{aligned}$$

where the last inequality follows from Claim 4.4. □

We can now use Claim 4.6 to show that the sum in (4) is not zero. Let $I^* \in \Gamma$ be the set with unique maximal d_{I^*} given by Claim 4.6. Rewrite (4) in the following form

$$\begin{aligned}\det(R \cdot D) &= \sum_{\substack{I \subset [n] \\ |I|=k}} \det(R_I) \cdot \det(D_I) \\ &= \sum_{I \in \Gamma} \det(R_I) \cdot \det(D_I) \\ &= \det(R_{I^*}) \cdot \det(D_{I^*}) + \sum_{\substack{I \in \Gamma \\ I \neq I^*}} \det(R_I) \cdot \det(D_I). \qquad (5)\end{aligned}$$

The degree of the first summand in (5) is at least

$$\deg\left(\det(R_{I^*}) \cdot \det(D_{I^*})\right) = d_{I^*} + \deg\left(\det(D_{I^*})\right) \geq d_{I^*}.$$

Using the 'moreover' part of Claim 4.6 we can upper bound the degrees of the other summands in (5). That is, for all $I \in \Gamma$ different from I^* we have

$$\deg\left(\det(R_I) \cdot \det(D_I)\right) = d_I + \deg\left(\det(D_I)\right) \leq d_I + dk < d_{I^*},$$

(we use the fact that all the entries of D are polynomials of degree at most d). Therefore, the sum in (5) cannot be zero. This concludes the proof of Theorem 4.1. □

5 Extractors for Polynomial Sources

In this section we describe our construction of an extractor for full rank polynomial sources and prove Theorem 2. As was mentioned in the introduction, this construction, together with the rank extractor constructed in previous sections, will give an extractor for polynomial sources of any rank. In order to describe our construction we require some additional notations. Let $\mathbb{F}$ be a field of prime cardinality p. For an integer $M \leq p$, we denote by $\mathrm{mod}_M : \mathbb{F} \to \{0, \ldots, M-1\}$ the modulo-M function. For a vector $x \in \mathbb{F}^n$ we apply the function $\mathrm{mod}_M(x)$ coordinate wise. The following theorem directly implies Theorem 2.

Theorem 5.1. *There exist absolute constants $C > 0$ and $c > 0$ such that the following holds: Let k, d be integers and let $\mathbb{F}$ be a field of prime cardinality $p > d^{Ck}$. Let $m > 0$ be an integer such that $m < c \cdot log(p)$, let $M = 2^m$ and define the function $E : \mathbb{F}^k \to \{0,1\}^{km}$ as $E(y) \triangleq mod_M(y)$. Then for every (k, k, d)-polynomial source Y over $\mathbb{F}$, the distribution $E(Y)$ is ϵ-close to uniform with $\epsilon = p^{-\Omega(1)}$.*

Notice that the construction of the extractor is very simple - taking a module in each coordinate. Proving that this is an extractor is much more complicated. The main tool in the proof of Theorem 5.1 will be a theorem of Bombieri [Bom66] giving an exponential sum estimate for low degree polynomials defined over curves (one dimensional varieties). We refer the reader to Appendix A for a discussion of the basic notions of algebraic geometry used in the proof.

5.1 Preliminaries for the proof of Theorem 5.1

5.1.1 Block Distributions

Our proof will rely on the following standard lemmas concerning block distributions.

Lemma 5.2. *Let A be some finite set and let $X = (X_1, \dots, X_k)$ be a distribution on A^k. Let $0 < \epsilon < 1$ and suppose that X_1 is ϵ-close to uniform. Suppose also that for each $2 \le i \le k$ there exists a set $S_i \subset A^{i-1}$ such that*

1. $\mathbf{Pr}[(X_1, \dots, X_{i-1}) \in S_i] \ge 1 - \epsilon$ *and*
2. *For each $s \in S_i$, the conditional distribution $(X_i | (X_1, \dots, X_{i-1}) = s)$ is ϵ-close to uniform.*

Then X is $O(k \cdot \epsilon)$-close to uniform.

Proof. We will prove the lemma for $k = 2$ (the general case will follow by a straight-forward induction). Let $T \subset A^2$ be some non empty set. It suffices to show that $\left|\Pr[(X_1, X_2) \in T] - |T|/|A|^2\right| \le O(\epsilon)$. For each $a \in A$ let $T_a = T \cap (\{a\} \times A)$. Let $S = S_2 \subset A$ be the set from the lemma. We have that

$$\begin{aligned}
\mathbf{Pr}[(X_1, X_2) \in T] &= \sum_{a \in A} \mathbf{Pr}[X_1 = a] \cdot \mathbf{Pr}[X_2 \in T_a | X_1 = a] \\
&\le \epsilon + \sum_{a \in S} \mathbf{Pr}[X_1 = a] \cdot \mathbf{Pr}[X_2 \in T_a | X_1 = a] \\
&\le 2\epsilon + \sum_{a \in S} \mathbf{Pr}[X_1 = a] \cdot \frac{|T_a|}{|A|} \\
&\le 3\epsilon + \sum_{a \in A} \frac{|T_a|}{|A|^2} = 3\epsilon + \frac{|T|}{|A|^2}.
\end{aligned}$$

Similarly, we can show an inequality in the opposite direction and so we conclude that (X_1, X_2) is 3ϵ-close to uniform. □

For our proof we require a modified version of this last lemma. In the modified version we fix not only the prefix of the distribution, but rather all indices except the i'th one. We recall our notation that for a vector $v = (v_1, \dots, v_n)$ and for an index $i \in [n]$ we have $v^{(-i)} = (v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n)$. In some places we will define a new vector of length $n - 1$ by writing $u = u^{(-i)} \in A^{n-1}$. This means that the indices of u go from 1 to n, skipping the i'th index. That is, $u = (u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n) \in A^{n-1}$.

Lemma 5.3. *Let A be some finite set and let $X = (X_1, \dots, X_k)$ be a distribution on A^k. Let $0 < \epsilon < 1$ and suppose that for each $1 \le i \le k$ there exists a set $S_i \subset A^{k-1}$ such that*

1. $\mathbf{Pr}[X^{(-i)} \in S_i] \ge 1 - \epsilon$ *and*
2. *For each $s^{(-i)} \in S_i$, the conditional distribution $(X_i | X^{(-i)} = s^{(-i)})$ is ϵ-close to uniform.*

Then X is $O(k \cdot \sqrt{\epsilon})$-close to uniform.

Proof. The lemma will follow by showing that X satisfies the conditions of Lemma 5.2 with ϵ replaced by $O(\sqrt{\epsilon})$. The first block X_1 (and indeed, all other blocks) is easily seen to be 2ϵ close to uniform by breaking it into a convex combination over all fixings of the other blocks, and throwing away those fixings not in S_1.

Now, let $i > 1$. For a prefix $(a_1, \ldots, a_{i-1}) \in A^{i-1}$ we define $P(a_1, \ldots, a_{i-1})$ to be the probability that $a^{(-i)} = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k)$ is in S_i when the additional elements $(a_{i+1}, \ldots, a_k)$ are chosen according the the distribution $(X_{i+1}, \ldots, X_k | X_1 = a_1, \ldots, X_{i-1} = a_{i-1})$. A simple averaging argument shows that the set $S'_i = \{(a_1, \ldots, a_{i-1}) \,|\, P(a_1, \ldots, a_{i-1}) \geq 1 - \sqrt{\epsilon}\}$ has probability at least $1 - \sqrt{\epsilon}$ in the distribution of $(X_1, \ldots, X_{i-1})$. We can thus, apply Lemma 5.2 with the sets S'_i and with ϵ replaced by $2\epsilon + \sqrt{\epsilon} = O(\sqrt{\epsilon})$. □

5.1.2 Distributions With Small Fourier Coefficients

The following lemma is an extension of the now folklore Vazirani XOR Lemma [Gol95] and is used [Bou07, BRSW06] to extract randomness from distributions with bounded Fourier coefficients. What the lemma says is that if we have a distribution X with a bound of $p^{-\Omega(1)}$ on all of its Fourier coefficients then we can deterministically extract from X (using the modulo function) $\Omega(\log(p))$ bits that are $p^{-\Omega(1)}$-close to uniform. The following formulation of the lemma follows from the version proved in [Rao07].

Lemma 5.4. *Let p be a prime number and let $0 < \alpha < 1$ be such that $\log(p) < p^{\alpha/2}$. Let X be a distribution on $\mathbb{F}$ - the field of p elements. Suppose that for every non-trivial additive character $\chi : \mathbb{F} \to \mathbb{C}^*$ we have the bound $\mathbb{E}[\chi(X)] \leq p^{-\alpha}$. Let $m = \lfloor (\alpha/2) \cdot \log(p) \rfloor$, let $M = 2^m$ and let $Y = mod_M(X)$ be an m-bit random variable. Then Y is $p^{-\alpha/4}$-close to uniform.*

5.1.3 Some Basic Facts on Varieties

In the proof of Theorem 5.1 we will use some facts regarding sets of the form $V = \{x \in \bar{\mathbb{F}}^n \,|\, f_i(x) = 0\,, i \in [r]\}$, where $\bar{\mathbb{F}}$ is the algebraic closure of a finite field and $f_1, \ldots, f_r$ are polynomials. These sets are algebraic varieties that are defined as an intersection of hypersurfaces. We include here three lemmas that will be used directly in the proof in the hope of making this section more readable. Readers less familiar with the notions of algebraic geometry are referred to the Appendix (or to any standard text on the subject) for the "bigger picture".

In the following $\bar{\mathbb{F}}$ denote the algebraic closure of a prime finite field $\mathbb{F}$. A variety is the set of common zeros of several polynomials. Intuitively speaking, an irreducible variety is a variety that is not the union of two or more distinct varieties. Every variety can be decomposed into a union of irreducible varieties and this decomposition is unique up to ordering. The notion of dimension used in the three lemmas below is defined formally in the appendix and can be thought of as a generalization of the same notion for affine subspaces. The next lemma, proved in the appendix, gives an upper bound on the number of irreducible components of a variety.

Lemma 5.5. *(Lemma A.31 in the Appendix) Let $f_1, \ldots, f_r \in \mathbb{F}[x_1, \ldots, x_n]$ be non-constant polynomials of degrees $d_1, \ldots, d_r$, respectively, and let $D = d_1 \cdots d_r$. Let $V = \{x \in \bar{\mathbb{F}}^n \,|\, f_i(x) = 0, i \in [r]\}$. Assume that V is non-empty and $dim(V) = n - r$. Then the number of irreducible components of V is at most D.*

The next lemma, used to prove Lemma 5.7, gives sufficient conditions under which the dimension of a variety, which is defined as the set of zeros of r polynomials in n variables, has dimension $n-r$. This lemma is proved in the Appendix.

Lemma 5.6. *(Lemma A.29 in the Appendix) Let $0 < r < n$ be integers and let $f_1, \ldots, f_r \in \mathbb{F}[x_1, \ldots, x_n]$ be non-constant polynomials. For each $i \in [r]$, let $H_i = \{x \in \bar{\mathbb{F}}^n \,|\, f_i(x) = 0\}$ and let $V_i = H_1 \cap \ldots \cap H_i$. Suppose that for each $2 \leq i \leq r$, f_i does not vanish identically on any of the irreducible components of the affine variety V_{i-1}. Then, if V_r is non-empty it is an affine variety all of whose irreducible components are of dimension $n-r$.*

Consider a system of $n-1$ polynomial equations in n variables. The next lemma gives a bound on the number of 'shifts' of the system for which the set of solutions has dimension larger than one (for the precise meaning of 'shift' see the lemma).

Lemma 5.7. *Let $\mathbb{F}$ be a finite field of size p and let $\bar{\mathbb{F}}$ denote its algebraic closure. Let $f_1, \ldots, f_{n-1} \in \mathbb{F}[x_1, \ldots, x_n]$ be polynomials of degree $\leq d$. For every $a = (a_1, \ldots, a_{n-1}) \in \mathbb{F}^{n-1}$ let $\hat{V}_a = \{x \in \bar{\mathbb{F}}^n \,|\, f_i(x) = a_i \,, i \in [n-1]\}$ and let $A = \{a \in \mathbb{F}^{n-1} \,|\, \hat{V}_a \neq \emptyset$ and $\dim(\hat{V}_a) \neq 1\}$. Then $|A| \leq nd^n p^{n-2}$.*

Proof. In order to bound $|A|$ we will describe an injective mapping from A to some small set. Fix some $a = (a_1, \ldots, a_{n-1}) \in A$. For $i \in [n-1]$ let $H_i = \{x \in \bar{\mathbb{F}}^n \,|\, f_i(x) = a_i\}$ be the hypersurface defined by the i'th restriction and let $U_i = H_1 \cap \ldots \cap H_i$ so that $U_{n-1} = \hat{V}_a$. Using Lemma 5.6 we see that if $\hat{V}_a$ is not empty and $\dim(\hat{V}_a) \neq 1$ then there must be some $2 \leq i \leq n-1$ such that H_i contains one of the irreducible components of U_{i-1}. Let i' be the smallest i satisfying this condition and let $0 < L \leq d^n$ be the index of the corresponding irreducible component of $U_{i'-1}$ (using some arbitrary ordering of the components of $U_{i'-1}$), where the bound of d^n on L follows from Lemma 5.5. Observe that if we are given the set of values $\{a^{(-i')}, i', L\}$ we can determine $a_{i'}$ and so recover a. Therefore, there exists an injective mapping from A into the set $\mathbb{F}^{n-2} \times [n] \times [d^n]$. Therefore $|A| \leq nd^n \cdot p^{n-2}$. □

5.1.4 A Theorem of Bombieri

The final ingredient we require for the proof of Theorem 5.1 is an exponential sum estimate due to Bombieri [Bom66]. We quote here a weak version of Bombieri's Theorem which is sufficient for our needs (we restate and derive this version of the theorem as Theorem A.37 in the appdendix).

Theorem 5.8 (Theorem 6 in [Bom66]). *Let p be a prime and let $1 < d$ be an integer such that $d^n < p$. Let $\mathbb{F}$ be the field of p elements and let $\bar{\mathbb{F}}$ be its algebraic closure. Let $f_1, \ldots, f_{n-1} \in \mathbb{F}[x_1, \ldots, x_n]$ be $n-1$ polynomials of degree $\leq d$ such that the set $\hat{V} = \{x \in \bar{\mathbb{F}}^n | f_1(x) = \ldots = f_{n-1}(x) = 0\}$ is a curve. Let $g \in \mathbb{F}[x_1, \ldots, x_n]$ be a polynomial of degree $\leq d$ that is non-constant on at least one of the irreducible components of $\hat{V}$. Let $\hat{V} = \hat{V}_1 \cup \ldots \cup \hat{V}_L$ be the decomposition of $\hat{V}$ into irreducible components. Let $\hat{U}$ be the union of those irreducible components of $\hat{V}$ on which $g(x)$ is non constant and let $U = \hat{U} \cap \mathbb{F}$. Let $\chi : \mathbb{F} \to \mathbb{C}^*$ be a non-trivial additive character of $\mathbb{F}$. Then*

$$\left| \sum_{x \in U} \chi(g(x)) \right| \leq 4d^{2n} \cdot p^{1/2}.$$

5.2 Proof of Theorem 5.1

Let $Y : \mathbb{F}^k \to \mathbb{F}^k$ be a (k, k, d)-polynomial source and let $f = (f_1, \dots, f_k) \in \mathbb{F}[x_1, \dots, x_k]$ be a vector of polynomials of degree at most d such that $Y(x) = f(x) = (f_1(x), \dots, f_k(x))$. For $i \in [k]$ and $a = a^{(-i)} \in \mathbb{F}^{k-1}$, we let $V_a = \left\{x \in \mathbb{F}^k \mid f^{(-i)}(x) = a\right\}$ and also $\hat{V}_a = \left\{x \in \bar{\mathbb{F}}^k \mid f^{(-i)}(x) = a\right\}$, where $\bar{\mathbb{F}}$ denotes the algebraic closure of $\mathbb{F}$. For a non trivial additive character $\chi : \mathbb{F} \to \mathbb{C}^*$, such that $V_a \neq \emptyset$ we define the exponential sum

$$\Upsilon_\chi(a) = \frac{1}{|V_a|} \sum_{x \in V_a} \chi(f_i(x)).$$

In view of Lemma 5.3 and Lemma 5.4 the theorem will follow from the following lemma.

Lemma 5.9. *Using the above notations, there exists* $0 < \alpha < 1$ *such that for every* $i \in [k]$ *there exists a set* $S_i \subset \mathbb{F}^{k-1}$ *such that*

1. $f^{(-i)}(x)$ *lands in* S_i *with probability at least* $1 - p^{-\alpha}$, *when* x *is chosen uniformly in* $\mathbb{F}^k$.
2. *For every* $a = a^{(-i)} \in S_i$ *and for every non trivial* χ, $|\Upsilon_\chi(a)| \leq p^{-\alpha}$.

Before proving the lemma we proceed to show how it is used to complete the proof of Theorem 5.1. Let us denote by

$$Z_i = \mathrm{mod}_M(f_i(x))$$

the random variable representing the i'th block of $E(Y)$. Let $0 < \alpha < 1$ be the constant given by Lemma 5.9. Let $i \in [k]$ and let $S_i \subset \mathbb{F}^{k-1}$ be the set given by Lemma 5.9.

We will define subsets $S'_i \subset [M]^{k-1}$ and then show that the distribution of $Z = E(Y)$ satisfies the conditions of Lemma 5.3 with the sets $S'_1, \dots, S'_k$ and with $\epsilon = p^{-\Omega(1)}$. The set S'_i will include all elements $b^{(-i)} \in [M]^{k-1}$ such that, when we condition on the event $Z^{(-i)} = b^{(-i)}$, we get that $f^{(-i)}(x)$ lands in S_i with probability at least $1 - p^{-\alpha/2}$. From Markov's inequality and from part (1) of Lemma 5.9 we have that

$$\Pr[b^{(-i)} \in S'_i] \geq 1 - p^{-\alpha/2}.$$

We now fix a specific value $b = b^{(-i)} \in S'_i$ and show that Z_i is close to uniform, even after we condition on the event $Z^{(-i)} = b^{(-i)}$. Denote by $Z_i(b)$ the distribution of Z_i conditioned on $Z^{(-i)} = b^{(-i)}$. Let $A \subset \mathbb{F}^{k-1}$ be the set of elements $a^{(-i)}$ that map to $b^{(-i)}$ by the function $\bmod_M(\cdot)$ and let $A' = A \cap S_i$. By the definition of S'_i we have that $Z_i(b)$ is $p^{-\alpha/2}$-close to a convex combination of distributions $W_i(a) = (Z_i | f^{(-i)}(x) = a)$ taken over all $a = a^{(-i)} \in A'$ (we simply throw away all elements $a \in A \setminus A'$ and add them to the error). We now use part (2) of Lemma 5.9 together with Lemma 5.4 to get that each $W_i(a)$ in the above convex combination is $p^{-\Omega(1)}$-close to uniform. Therefore $Z_i(b)$ is also $p^{-\Omega(1)}$-close to uniform. We have proved that $Z = (Z_1, \dots, Z_k)$ satisfies all the conditions of Lemma 5.3 with $\epsilon = p^{-\Omega(1)}$ and so we are done since $O(k \cdot \sqrt{p^{-\Omega(1)}}) = p^{-\Omega(1)}$ when $p > d^{Ck}$ and C is sufficiently large.

5.2.1 Proof of Lemma 5.9

Let $i \in [k]$. We would like to distinguish between "good" and "bad" fixings of $f^{(-i)}(x)$. The "good" fixings will be those values $a = a^{(-i)} \in \mathbb{F}^{k-1}$ for which we can bound the exponential sum $\Upsilon_\chi(a)$. Before proving the Lemma formally let us describe briefly the intuition behind the proof. Each fixing $f^{(-i)}(x) = a^{(-i)}$ defines a variety V. We would like to apply Bombieri's Theorem to bound the exponential sum of $f_i(x)$ over this variety. In order to do so we need to make sure that V is a curve and that $f_i(x)$ is not constant on 'enough' of the components of the curve V (where the word 'enough' takes into account the number of points in $\mathbb{F}$ in each component). The fact that most fixings satisfy the first condition, that V is a curve, will follow from a counting argument, based on a version of Bezout's theorem. The second condition will follow from Wooley's Theorem (Theorem 2.8). Intuitively, Wooley's theorem tells us that the image of f is close to having high min-entropy. Clearly, this should allow us to bound the size of those components on which $f_i(x)$ is constant (for 'most' fixings of $f^{(-i)}(x)$).

In order to be able to define these "good" fixings of $f^{(-i)}(x)$ we need to consider the singular points of the mapping $f(x)$, namely the zeros of its Jacobian. Let $J(x) = \det\left(\frac{\partial f}{\partial x}\right)$ be the determinant of the Jacobian of $f(x)$, which is a non zero polynomial since the source Y has full rank. Let $\mathrm{Sing} = \{x \in \mathbb{F}^k \mid J(x) = 0\}$ be the set of singular points and for each $a = a^{(-i)} \in \mathbb{F}^{k-1}$ let $\mathrm{Sing}_a = \mathrm{Sing} \cap V_a$.

Definition 5.10. *We say that $a = a^{(-i)} \in \mathbb{F}^{k-1}$ is "good" if it satisfies the following three conditions:*

1. $|V_a| \geq p^{5/6}$.
2. $|Sing_a| \leq p^{1/6}$.
3. $\hat{V}_a$ *is a curve. That is,* $\dim(\hat{V}_a) = 1$.

We define the set $S_i \subset \mathbb{F}^{k-1}$ to be the set of all "good" a's.

The next claim shows that most a's are "good". Thus proving part (1) of Lemma 5.9.

Claim 5.11. *Let S_i be as above. Then $\mathbf{Pr}[f^{(-i)} \in S_i] \geq 1 - p^{-\Omega(1)}$, where the probability is over uniformly chosen $x \in \mathbb{F}^k$.*

Proof. Let $a = a^{(-i)} \in \mathbb{F}^{k-1}$ be the random variable sampled by $a = f^{(-i)}(x)$, x uniform. For $1 \leq j \leq 3$ let E_j denote the event that a satisfies condition j in Definition 5.10. We can write

$$\mathbf{Pr}[a \text{ is "bad"}] \leq \mathbf{Pr}[E_1^c] + \mathbf{Pr}[E_2^c] + \mathbf{Pr}[E_1 \wedge E_2 \wedge E_3^c]. \tag{6}$$

We will bound each of these three probabilities independently by $p^{-\Omega(1)}$, which will prove the claim. The first probability can be seen to be bounded by $p^{-1/6}$ by a simple union bound on all a's with small $|V_a|$.

To bound the second probability we first observe that $|\mathrm{Sing}| \leq \deg(J(x)) \cdot p^{k-1} \leq dk \cdot p^{k-1}$. Therefore, the number of different a's not satisfying condition (2) is at most $dk \cdot p^{k-7/6}$. From Theorem 2.8 we have that for every $a = a^{(-i)} \in \mathbb{F}^{k-1}$ the set V_a contains at most $d^k \cdot p$ non-singular points. Therefore, the size of the union of all V_a's for which condition (2) is not satisfied is bounded by

$$kd \cdot p^{k-1} + (kd \cdot p^{k-7/6})(d^k \cdot p) \leq p^{k-\Omega(1)}$$

(the first term counts all singular points and the second term counts all non singular points), where the inequality holds for $p > d^{Ck}$ for sufficiently large constant C. Therefore the second probability in Eq. 6 is also bounded by $p^{-\Omega(1)}$.

We now bound the third probability in Eq. 6. Let $A \subset \mathbb{F}^{k-1}$ be the set of a's satisfying conditions (1) and (2) but not (3) in the definition of a "good" a. We first observe that Lemma 5.7 gives us the bound $|A| \leq kd^k \cdot p^{k-2}$ on the size of A. Now, For each $a \in A$ the size of V_a is bounded by $p^{1/6} + d^k \cdot p$ (V_a does not contain many singular points since a satisfies condition (2)). Therefore, we have that

$$\sum_{a \in A} |V_a| \leq |A| \cdot (p^{1/6} + d^k \cdot p) \leq kd^k \cdot p^{k-2} \cdot (p^{1/6} + d^k \cdot p) \leq p^{k-\Omega(1)},$$

(when $p > d^{Ck}$ and C is sufficiently large). This completes the proof of the claim. □

We now move to proving part (2) of Lemma 5.9. We will show that for every $a = a^{(-i)} \in S_i$ and for every non trivial character χ the sum $|\Upsilon_\chi(a)|$ is bounded by $p^{-\Omega(1)}$.

Claim 5.12. *Let $a = a^{(-i)} \in S_i$. Then we have the bound $|\Upsilon_\chi(a)| \leq p^{-\Omega(1)}$.*

Proof. Let $\hat{V}_a = \hat{C}_1 \cup \ldots \cup \hat{C}_L$ be the decomposition of the curve $\hat{V}_a$ into irreducible components and let $C_j = \hat{C}_j \cap \mathbb{F}^k$ for $j \in [L]$. From Lemma 5.5 we have that $L \leq d^k$. We wish to use Theorem 5.8 to bound $|\Upsilon_\chi(a)|$. Our first step will be to show that the polynomial $f_i(x)$ can be constant only on those irreducible components $\hat{C}_j$ that have few points in $\mathbb{F}_p$. To show this, notice that if the polynomial $f_i(x)$ is constant on one of the irreducible components $\hat{C}_j$ then , using Theorem 2.8 and part (2) of the definition of "good" a's, we get that $|C_j| \leq p^{1/6} + d^k$.

We now consider the modified curve $\hat{U}_a$ constructed by taking the union of those components $\hat{C}_j$ of $\hat{V}_a$ for which $|C_j| > p^{1/6} + d^k$ and let $U_a = \hat{U}_a \cap \mathbb{F}^k$. We can now use Theorem 5.8 to get the bound

$$\left| \sum_{x \in U_a} \chi(f_i(x)) \right| \leq 4d^{2k} \cdot p^{1/2},$$

which translates into the bound

$$\left| \sum_{x \in V_a} \chi(f_i(x)) \right| \leq d^k \cdot (p^{1/6} + d^k) + 4d^{2k} p^{1/2} \leq p^{2/3}$$

(separating the sum into points in the small components and in the large components) where the inequality holds when $p > d^{Ck}$, C sufficiently large. Dividing this sum by $|V_a| > p^{5/6}$ we get the required bound of $p^{-\Omega(1)}$ on $|\Upsilon_\chi(a)|$. □

Combining the above two claims concludes the proof of Lemma 5.9. □

6 Improving the Output Length

The extractor constructed in Section 5 can extract a constant fraction of the min-entropy of the source. It was suggested to us by Salil Vadhan that we can extract almost all of the min-entropy by using special properties of the source. This indeed works, and in this section we explain how.

We recall the notations of the last section: let $Y : \mathbb{F}^k \to \mathbb{F}^k$ be a (k, k, d)-polynomial source. Before describing the improved construction we need to define *seeded extractors*. For this section only we denote by U_s the uniform distribution on s bits.

Definition 6.1. *A function $E : \{0,1\}^n \times \{0,1\}^s \to \{0,1\}^m$ is an (r,ϵ)-*seeded extractor *if for every distribution X such that $H_\infty(X) \geq r$ the distribution $E(X, U_s)$ is ϵ-close to uniform. E is said to be* explicit *if it can be computed in polynomial time.*

Roughly speaking the method to extract many bits from Y is as follows: Let $E_1 : \mathbb{F} \to \{0,1\}^{m_1}$ be the extractor for distributions with small Fourier coefficients given by Lemma 5.4 (namely the mod 2^{m_1} function) and let $E_2 : \mathbb{F}^{k-1} \times \{0,1\}^s \to \{0,1\}^{m_2}$ be any seeded extractor with seed length s and output length m_2. Consider the composition of these two extractors given by $E(Y) = E_2(Y^{(-k)}, E_1(Y_k))$ (recall that $Y^{(-k)} = (Y_1, \ldots, Y_{k-1})$) in which the role of the uniform seed is taken by $E_1(Y_k)$. We would like to claim that $E(Y)$ is close to uniform. The first thing to observe is that m_1 has to be larger than s. This requirement will be easy to satisfy since in our setting, when $p \geq d^{O(k)}$, the output of E_1 will be larger then the seed length of standard seeded extractors. The more important thing to justify is the fact that we can replace the uniform seed of E_2 with a seed that is correlated with the source - $Y^{(-k)}$. This can be done since for 'most' fixings of $Y^{(-k)}$, the random variable $E_1(Y_k)$ is close to uniform (this follows from Bombieri's Theorem and the analysis of Section 5). We formalize this intuition in the following theorem:

Theorem 6.2. *Let k, d be integers and let $\mathbb{F}$ be a prime field of size $p > d^{\Omega(k)}$. Let $m_1 = c \cdot \log(p)$ for some small absolute constant c. Let $E_1 : \mathbb{F} \to \{0,1\}^{m_1}$ be the function $E_1(j) = \mod_{2^{m_1}}(j)$ and let $E_2 : \mathbb{F}^{k-1} \times \{0,1\}^s \to \{0,1\}^{m_2}$ be an (r,ϵ)-seeded extractor.*[2] *Suppose that $m_1 \geq s$ and $r \leq (k-1) \cdot \log\left(\frac{p}{2d}\right)$. Then for any (k,k,d)-polynomial source $Y : \mathbb{F}^k \to \mathbb{F}^k$ we have that $E_2(Y^{(-k)}, E_1(Y_k))$ is ϵ'-close to uniform, with $\epsilon' = \epsilon + p^{-\Omega(1)}$ (we will use the convention that if $m_1 > s$ then E_2 uses only the first s bits of $E_1(Y_k)$).*

Proof. Assume w.l.o.g that $m_1 = s$. Using Lemma 5.9 together with Lemma 5.4 we get that with probability at least $1 - p^{-\Omega(1)}$ over a random fixing $Y^{(-k)} = b^{(-k)}$, the distribution $\left(E_1(Y_k) | Y^{(-k)} = b^{(-k)}\right)$ is $p^{-\Omega(1)}$-close to uniform. This means that the joint distribution $(Y^{(-k)}, E_1(Y_k))$ is $p^{-\Omega(1)}$-close to $(Y^{(-k)}, U_s)$. Therefore, we have that $E_2(Y^{(-k)}, E_1(Y_k))$ is $p^{-\Omega(1)}$-close to $E_2(Y^{(-k)}, U_s)$ which is $\epsilon + p^{-\Omega(1)}$ close to uniform by the properties of E_2. Here we use the fact that $r \leq (k-1) \cdot \log\left(\frac{p}{2d}\right)$ and that, from Lemma 2.9, $Y^{(-k)}$ is $p^{-\Omega(1)}$-close to having min-entropy at least $(k-1) \cdot \log\left(\frac{p}{2d}\right)$. □

Applying the last theorem with an appropriate seeded extractor enables us to construct a deterministic extractor for polynomial sources that extract any constant fraction of the entropy of the source. It is possible to increase further the output length by using different seeded extractors. However, using current state-of-the-art seeded extractors, this would cost in terms of the error of the final construction. In order to avoid these complications we concentrate on extracting only a constant fraction (arbitrarily close to 1) of the min entropy.

Theorem 6.3. *Let k and $d > 1$ be integers and let $\mathbb{F}$ be a field of prime cardinality $p > d^{\Omega(k)}$. Let $0 < \alpha < 1$. Then, there exists a function $E : \mathbb{F}^k \to \{0,1\}^m$ that is an explicit (k,d,ϵ)-extractor for polynomial sources over $\mathbb{F}^k$ with $m = (1-\alpha) \cdot k \cdot \log\left(\frac{p}{2d}\right)$ and $\epsilon = p^{-\Omega(1)}$.*

Proof. We use the seeded extractors of [RRV99] in conjunction with Theorem 6.2. In [RRV99] it is shown that there exists an explicit (r,ϵ)-seeded extractor $E_2 : \mathbb{F}^{k-1} \times \{0,1\}^s \to \{0,1\}^{m_2}$ with the

[2] We can safely ignore the technicality that p^{k-1} is not a power of two.

following parameters:

$$r = \lfloor (k-1) \cdot \log\left(\frac{p}{2d}\right) \rfloor,$$

$$\epsilon = p^{-\Omega(1)},$$

$$m_2 \geq (1 - \alpha/2) \cdot r$$

$$s = O(\log^2(k \cdot \log(p)) + \log(1/\epsilon)) = O(\log(p)).$$

Plugging E_2 into the setting described in Theorem 6.2 we get an extractor with output length $m_2 \geq (1-\alpha/2)(k-1) \cdot \log\left(\frac{p}{2d}\right)$ which is larger then $(1-\alpha) \cdot k \cdot \log\left(\frac{p}{2d}\right)$. □

7 Extractors For Weak Polynomial Sources

In this section we discuss the more general class of sources defined in the introduction as (n,k,d)-weak polynomial sources. Our final goal will be to prove Theorem 4, which we restate here for convenience:

Theorem 4. *There exists absolute constants C and c such that the following holds: Let $k \leq n$ and $d > 1$ be integers and let $d' = 8k^2d^3n$. Let $\mathbb{F}$ be a field of prime cardinality $p > (d')^{Ck}$. Then, there exists a function $E : \mathbb{F}^n \to \{0,1\}^m$ that is an explicit (k,d,ϵ)-extractor for weak polynomial sources over $\mathbb{F}^n$ with $m = \lfloor c \cdot k \cdot \log(p) \rfloor$ and $\epsilon = p^{-\Omega(1)}$.*

Theorem 4 will be a simple corollary of the following theorem, which shows that any (n,k,d)-WPS is close to a convex combination of (n,k,d)-polynomial sources.

Theorem 7.1. *Let $\mathbb{F}$ be a field of prime cardinality p. Let $k \leq n$ and d be integers such that $p > \max\{4D^2, 2^{10}\}$, where $D = (2k+1)d^{2k}$. Let X be an (n,k,d)-WPS over $\mathbb{F}$. Then X is δ-close to a convex combination of (n,k,d)-polynomial sources over $\mathbb{F}$, with $\delta = \frac{d \cdot k}{p}$.*

Before proving Theorem 7.1 we show how it can be used to prove Theorem 4.

Proof of Theorem 4. Let X be an (n,k,d)-WPS. We take the extractor $E : \mathbb{F}^k \to \{0,1\}^m$ to be the one given by Corollary 1.4 (namely, the extractor for polynomial sources). Using Theorem 7.1 we get that X is δ-close to a convex combination of (n,k,d)-polynomial sources, with $\delta = \frac{d \cdot k}{p} = p^{-\Omega(1)}$ (when $p > (d')^{Ck}$ and C sufficiently large). We know from Corollary 1.4 that E is a (k,d,ϵ)-extractor for polynomial sources over $\mathbb{F}^n$, with $\epsilon = p^{-\Omega(1)}$. Therefore, $E(X)$ is δ-close to a convex combination of distributions, each of which is ϵ-close to uniform. It follows, using standard probability theory, that $E(X)$ is $(\delta + \epsilon) = p^{-\Omega(1)}$-close to uniform.

7.1 Proof of Theorem 7.1

The proof of the theorem will be in two steps. The first step will be to show that every (n,k,d)-WPS is sampled by a mapping $x : \mathbb{F}^n \to \mathbb{F}^n$ such that $\text{rank}(x) \geq k$. The second step will be to show that a distribution sampled by such a mapping is close to a convex combination of (n,k,d)-polynomial sources. The first step of the proof of Theorem 7.1 is given by the following lemma.

Lemma 7.2. *Let $\mathbb{F}$ be a field of prime cardinality p. Let $k \leq n$ and d be integers such that $p \geq \max\{4D^2, 2^{10}\}$, where $D = (2k+1) \cdot d^{2k}$. Let X be an (n,k,d)-WPS over $\mathbb{F}$. Then there exists a mapping $x \in \mathcal{M}(\mathbb{F}^n \to \mathbb{F}^n)$ with rank $\geq k$ such that $X = x(U_n)$.*

The main thing that is needed in order to prove Lemma 7.2 is to show that if a polynomially sampled distribution has high entropy, then its rank is also high. In other words, we need to show that if the rank is low, so is the entropy. We achieve this kind of bound in two parts. The first part bounds the entropy of the output distribution of k *dependent* polynomials. That is, of k polynomials with rank at most $k-1$. This can be viewed as the 'base case' for the proof of Lemma 7.2.

Lemma 7.3. *Let $\mathbb{F}$ be a field of prime cardinality p. Let k, n and d be integers such that $p > D$, where $D = (n+1)d^n$. Let $f_1, \ldots, f_k \in \mathbb{F}[x_1, \ldots, x_n]$ be k algebraically dependent polynomials of total degree at most d. Let P denote the distribution of the mapping $f = (f_1, \ldots, f_k) : \mathbb{F}^n \to \mathbb{F}^k$ on a uniformly chosen input in $\mathbb{F}^n$. Then P has support size at most $D \cdot p^{k-1}$.*

Proof. From Theorem 3.3 we know that there exists a non zero polynomial $h \in \mathbb{F}[z_1, \ldots, z_k]$ of degree $\leq D$ such that $h(f_1(x), \ldots, f_k(x)) \equiv 0$ (notice that we use Theorem 3.3 with the roles of k and n reversed). Therefore, the support of P is contained in the zero set of h, whose size is bounded by $D \cdot p^{k-1}$ by Schwartz-Zippel (Lemma 2.6). □

The second auxiliary lemma we will need in the proof of Lemma 7.2 is the following lemma which will enable us to reduce the number of variables of a mapping (assuming the number of variables is considerably larger than the number of outputs) while maintaining both the rank and the overall entropy of the mapping.

Lemma 7.4. *Let $\mathbb{F}$ be a finite field of cardinality q. Let d, k, n, m be integers such that $2k \leq n$. Let $x \in \mathcal{M}(\mathbb{F}^n \to \mathbb{F}^m, d)$ be such that $H_\infty(x(U_n)) \geq k \cdot \log(q)$. Then, there exists an affine subspace $V \subset \mathbb{F}^n$ of dimension $2k$ such that the restriction of x to V has min entropy at least $k \cdot \log(q) - 2$. That is, if we denote by U_V the uniform distribution on V, then we have $H_\infty(x(U_V)) \geq k \cdot \log(q) - 2$.*

Proof. Take V to be a random affine subspace of dimension $2k$. For each $y \in \mathbb{F}^m$ let $S_y \triangleq \{t \in \mathbb{F}^n \mid x(t) = y\}$ and denote $r_y \triangleq |S_y| \cdot q^{-n} = \Pr[x(U_n) = y]$. Fix some $y \in \mathbb{F}^m$. The expectation, over the choice of V, of $|S_y \cap V|$ is $q^{2k} \cdot r_y$. We can also bound the variance of $|S_y \cap V|$ (using pairwise independence of the points on V) by $|S_y| q^{2k-n}(1 - q^{2k-n}) \leq q^{2k} \cdot r_y$. Applying Chebyshev's inequality, and using the fact that for all $y \in F^m$ we have $r_y \leq q^{-k}$, one can show that

$$\Pr_V[\,|S_y \cap V| > 4q^k\,] \leq \frac{r_y}{9}. \tag{7}$$

Using the union bound we get that the probability that there exists a y for which the event in (7) happens is bounded by $1/9$ and so there exists V such that for all $y \in F^m$ we have $|S_y \cap V| \leq 4q^k$. This completes the proof of the lemma since

$$\Pr[x(U_V) = y] = \frac{|S_y \cap V|}{q^{2k}} \leq 4q^{-k}.$$

□

The third auxiliary lemma we will use in the proof of Lemma 7.2 is the following lemma which enables us to reduce the number of polynomials from n to k while maintaining most of the entropy.

Lemma 7.5. *Let $\mathbb{F}$ be a finite field of cardinality q. Let $k \leq n$ be integers and let $0 < s \leq k$ be a real number. Let X be a distribution over $\mathbb{F}^n$ such that $H_\infty(X) \geq s \cdot \log(q)$. Then there exists a linear* *mapping $l : \mathbb{F}^n \to \mathbb{F}^k$ such that for every $\alpha > 0$ the distribution $l(X)$ is ϵ-close to having min entropy $\geq (s - \alpha) \cdot \log(q)$, where $\epsilon = \sqrt{2} \cdot q^{-\alpha/2}$.*

Proof. Let $\mathcal{L}$ denote the set of all linear mappings from $\mathbb{F}^n$ to $\mathbb{F}^k$ and let L be a random variable uniformly distributed over $\mathcal{L}$. Let us observe the average collision probability of $l(X)$ when we average over all $l \in \mathcal{L}$.

$$\begin{aligned}
\frac{1}{|\mathcal{L}|} \sum_{l \in \mathcal{L}} \mathrm{cp}(l(X)) &= \sum_{l \in \mathcal{L}} \Pr[L = l] \cdot \Pr_{x_1, x_2 \leftarrow X} [L(x_1) = L(x_2) \,|\, L = l] \\
&= \Pr_{x_1, x_2 \leftarrow X} [L(x_1) = L(x_2)] \\
&\leq \Pr_{x_1, x_2 \leftarrow X} [x_1 = x_2] + \Pr_{x_1, x_2 \leftarrow X} [L(x_1) = L(x_2) \,|\, x_1 \neq x_2] \\
&\leq q^{-s} \quad + \quad q^{-k} \leq 2q^{-s},
\end{aligned}$$

where in the last inequality we used the fact that the min entropy of X is at least $\log(q^s)$ and so $\mathrm{cp}(X) \leq q^{-s}$. Therefore, there exists $l \in \mathcal{L}$ such that $\mathrm{cp}(l(X)) \leq 2q^{-s}$. Let $\alpha > 0$ and let us use Lemma 2.4 with $a = \frac{q^\alpha}{2}$ and $b = q^{s-\alpha}$. We therefore have $\mathrm{cp}(l(X)) \leq \frac{1}{ab}$ and so, by the lemma, $l(X)$ is $(1/\sqrt{a})$-close to having min entropy at leat $\log(b) = (s - \alpha) \cdot \log(q)$. □

One more simple auxiliary claim we will require is the following claim.

Claim 7.6. *Let $0 < \epsilon < 1/4$. Let X be some distribution on some finite set Γ. Suppose that X is ϵ-close to a distribution with support size at most M. Then X is $1/4$-far from any distribution with min entropy at least* $\log(2M)$.

Proof. Assume towards a contradiction that there exists a distribution Y on Γ such that $\mathrm{H}_\infty(Y) \geq \log(2M)$ and $X \overset{\delta}{\sim} Y$ with $\delta \leq 1/4$. From the assumption on X we know that there exists a set $A \subset \Gamma$ with $|A| \leq M$ such that $\mathbf{Pr}[X \in A] \geq 1 - \epsilon$. We therefore have that $\mathbf{Pr}[Y \in A] \geq 1 - \epsilon - \delta > 1/2$. Therefore, since $\mathbf{Pr}[Y = a] \leq 2^{-\log \mathrm{H}_\infty(Y)} \leq \frac{1}{2M}$, we get that $\mathbf{Pr}[Y \in A] \leq |A| \cdot \frac{1}{2M} \leq 1/2$, a contradiction. □

We are now ready to prove Lemma 7.2.

Proof of Lemma 7.2 Let $x = x(t) \in \mathcal{M}(\mathbb{F}^n \to \mathbb{F}^n, d)$ be a mapping such that $X = x(U_n)$. We will show that $\mathrm{rank}(x) \geq k$. Assume towards a contradiction that $\mathrm{rank}(x) < k$. Using Lemma 7.4 we can replace x with a new polynomial mapping $\tilde{x} \in \mathcal{M}(\mathbb{F}^m \to \mathbb{F}^n, d)$,with $m = \min(n, 2k)$, and such that (a) $rank(\tilde{x}) \leq rank(x) < k$ and (b) $\mathrm{H}_\infty(\tilde{x}(U_m)) \geq (k - 1/4)\log(q)$. Let $\tilde{X}$ denote the output distribution of $\tilde{x}$.

Next, we use Lemma 7.5 with parameters $\alpha = 1/4$ and $s = k - 1/4$. We get that there exists a linear mapping $l : \mathbb{F}^n \to \mathbb{F}^k$ such that $l(\tilde{X})$ is ϵ-close to having min-entropy at least $(k - 1/2) \cdot \log(p)$, where

$$\epsilon = \sqrt{2} \cdot p^{1/8} < 1/4,$$

where the last inequality uses the fact that $p > 2^{10}$.

Notice that the distribution $l(\tilde{X})$ is the output distribution of k *dependent* polynomials. To see this write $D = \frac{\partial x}{\partial t}$ and let A_l be a $k \times n$ matrix representing l. The partial derivative matrix of $l \circ x$

is simply $A_l \cdot D$ and the rank of this matrix is at most the rank of D, which we assumed is bounded by $k-1$. Theorem 3.3 now implies that the polynomials sampling $l(\tilde{X})$ are dependent.

We can now use Lemma 7.3 to get that $l(\tilde{X})$ has support size at most $D \cdot p^{k-1}$, where $D = (m+1)d^m$. Therefore, by Claim 7.6, $l(\tilde{X})$ is (1/4)-far from any distribution with min entropy at least $\log(2D \cdot p^{k-1})$. This implies

$$p^{k-1/2} < 2D \cdot p^{k-1},$$

which gives $p < 4D^2$, a contradiction. □

The second step in the proof of Theorem 7.1 is the following lemma.

Lemma 7.7. *Let $\mathbb{F}$ be a finite field. Let $k \leq n$ and d be integers. Let $x \in \mathcal{M}(\mathbb{F}^n \to \mathbb{F}^n, d)$ be a mapping with rank k. Let X be the distribution $x(U_n)$. Then X is ϵ-close to a convex combination of (n,k,d)-polynomial sources over $\mathbb{F}$, where $\epsilon = \frac{d \cdot k}{|\mathbb{F}|}$.*

Proof. Denote by D the sub-matrix of the first k rows and k columns of $\frac{\partial x}{\partial t}$, i.e.,

$$D = \begin{pmatrix} \frac{\partial x_1}{\partial t_1} & \cdots & \frac{\partial x_1}{\partial t_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_k}{\partial t_1} & \cdots & \frac{\partial x_k}{\partial t_k} \end{pmatrix}.$$

We can assume w.l.o.g that D is non singular (this can be obtained by relabelling the t's and x's). Let $f : \mathbb{F}^n \to \mathbb{F}$ be defined as $f(t) \triangleq \det(D)(t)$. By assumption, f is non-zero and $deg(f) \leq d \cdot k$. For $c = (c_{k+1}, \ldots, c_n) \in \mathbb{F}^{n-k}$ define the mapping $x_c : \mathbb{F}^k \to \mathbb{F}^n$ as x restricted to c, that is $x_c(t_1, \ldots, t_k) \triangleq x(t_1, \ldots, t_k, c_{r+1}, \ldots, c_n)$. Note that, the first k rows of $\frac{\partial x_c}{\partial t}$ are exactly D under the restriction $t_{k+1} = c_{k+1}, \ldots, t_n = c_n$. Thus $\frac{\partial x_c}{\partial t}$ has full rank whenever $f_c(t_1, \ldots, t_k) \triangleq f(t_1, \ldots, t_k, c_{k+1}, \ldots, c_n)$ is non-zero. Using Claim 2.7, $f_c \equiv 0$ with probability at most $\frac{d \cdot k}{|\mathbb{F}|}$ (for uniformly chosen c). Let X_c be the distribution $x_c(U_k)$. Then X is a convex combination of the X_c's. Moreover, using Lemma 2.1, X is $\frac{d \cdot k}{|\mathbb{F}|}$-close to a convex combination of the X_c's for which f_c is non-zero, and these X_c's are (n,k,d)-polynomial sources over $\mathbb{F}$. □

Proof of Theorem 7.1 We first apply Lemma 7.2 to get that X is sampled by a rank k mapping $x : \mathbb{F}^n \to \mathbb{F}^n$. Then we use Lemma 7.7 to show that $X = x(U_n)$ is δ-close to a convex combination of (n,k,d)-polynomial sources with $\delta = \frac{d \cdot k}{p}$. □

7.2 The Entropy of a Polynomial Mapping

We can use the results of the last section to show that, over sufficiently large fields, the distribution sampled by a low degree mapping $x \in \mathcal{M}(\mathbb{F}^n \to \mathbb{F}^n, d)$ is always close to having entropy approximately $k \cdot \log(p)$, where k is equal to the rank of x. This is a generalization of the affine case, where the entropy is *exactly* $k \cdot \log(p)$. This is stated formally by the following theorem.

Theorem 7.8. *Let $k \leq n$ and d be integers. Let $D = (2k+1)d^{2k}$ and let $0 < \delta < 1$ be a real number. Let $\mathbb{F}$ be a field of prime cardinality p such that $p > \max\{(2d)^{\frac{k}{\delta}}, 2^{\frac{10}{\delta}}, (2D)^{\frac{2}{\delta}}\}$. Let $x \in \mathcal{M}(\mathbb{F}^n \to \mathbb{F}^n, d)$ be of rank k and let $X = x(U_n)$ be the distribution sampled by x. Then*

1. *X has min entropy $\leq (k+\delta)\cdot\log(p)$.*

2. *X is ϵ-close to having min entropy $\geq (k-\delta)\cdot\log(p)$, where $\epsilon = \frac{2\cdot d\cdot k}{p}$.*

Proof. We start with a proof of 2, which is easier. We apply Lemma 7.7 to get that X is $\frac{d\cdot k}{p}$-close to a convex combination of (n,k,d)-polynomials sources. From Theorem 2.9 we have that every distribution in this convex combination is $\frac{d\cdot k}{p}$-close to having min entropy $\geq k\cdot\log\left(\frac{p}{2d}\right)$. It follows that X is $\frac{2\cdot d\cdot k}{p}$-close to having min entropy at least

$$k\cdot\log\left(\frac{p}{2d}\right) \geq (k-\delta)\cdot\log(p),$$

where the inequality follows from the bound $p \geq (2d)^{\frac{k}{\delta}}$.

We proceed to prove part 1 of the theorem. We can assume w.l.o.g that $k < n$, for otherwise an entropy upper bound of $n\cdot\log(p)$ would be trivial. Suppose for contradiction that $\mathrm{H}_\infty(x) > (k+\delta)\cdot\log(p)$. Using Lemma 7.4 we can replace x with a new polynomial mapping $\tilde{x}\in\mathcal{M}(\mathbb{F}^m\to\mathbb{F}^n,d)$,with $m=\min(n,2k)$, and such that (a) $rank(\tilde{x}) \leq rank(x) = k$ and (b) $\mathrm{H}_\infty(\tilde{x}(U_m)) \geq (k+\frac{3}{4}\delta)\log(p)$, where we need to use the following inequality

$$(k+\frac{3}{4}\delta)\log(p) \leq (k+\delta)\log(p) - 2,$$

which holds for $p > 2^{\frac{10}{\delta}}$.

Let $\tilde{X}$ denote the output distribution of $\tilde{x}$. We apply Lemma 7.5 with $\alpha = \delta/4$ to find a linear mapping $l:\mathbb{F}^n\to\mathbb{F}^{k+1}$ such that $l(\tilde{X})$ is ϵ'-close to having min-entropy at least $(k+\delta/2)\cdot\log(p)$ with $\epsilon' = \sqrt{2}\cdot p^{-\delta/8} < 1/4$ (here we use again the bound $p > 2^{\frac{10}{\delta}}$).

We proceed in a similar manner as in the proof of Lemma 7.2: We first use Lemma 7.3 to claim that $l(\tilde{X})$ has support size at most $D\cdot p^k$, where $D=(m+1)d^m$ (again, using the fact that $l\circ\tilde{x}$ has rank at most $\mathrm{rank}(\tilde{x})\leq k$). From this fact and from Claim 7.6 we deduce that

$$(k+\delta/2)\cdot\log(p) \leq \log(2D\cdot p^k),$$

which is a contradiction since $p > (2D)^{\frac{2}{\delta}}$. □

8 Rank Extractors Over The Complex Numbers

In this section we discuss the interpretation of rank extractors over the complex numbers. This interpretation will follow from the results appearing in [ER93], on algebraic independence and full-rank mappings over $\mathbb{C}$. The following theorem shows that over the complex numbers algebraic independence is equivalent to full rank.

Theorem 8.1. *[Theorem 2.3 in [ER93]] Let $x\in\mathcal{M}(\mathbb{C}^k\to\mathbb{C}^r,d)$ where $r\leq k$. The mapping x has full rank, that is, rank r, if and only if $x_1,\ldots,x_r$ are algebraically independent.*

The next theorem shows that for a mapping $x\in\mathcal{M}(\mathbb{C}^k\to\mathbb{C}^k,d)$, full rank is equivalent to having an image that is "essentially all" of $\mathbb{C}^k$. More precisely, all of $\mathbb{C}^k$ except for a set of measure zero. The theorem follows immediately from Theorem 2.4 in [ER93].

Theorem 8.2. *Fix any integers d, k and any $x \in \mathcal{M}(\mathbb{C}^k \to \mathbb{C}^k, d)$. The mapping x has full rank if and only if the image $x(\mathbb{C}^k)$ of x contains all of $\mathbb{C}^k$ except a set $Z \subseteq \mathbb{C}^k$ of measure zero in $\mathbb{C}^k$.*

Proof. Assume that x has full rank. In the proof of Theorem 2.4 in [ER93], it is shown that $x(\mathbb{C}^k)$ contains all of $\mathbb{C}^k$ except the set Z of zeros of some polynomial $H : \mathbb{C}^k \to \mathbb{C}$. Such a set Z has measure zero. Now assume $x(\mathbb{C}^k)$ contains all of $\mathbb{C}^k$ except for a set of measure zero in $\mathbb{C}^k$. Then $x(\mathbb{C}^k)$ is dense in $\mathbb{C}^k$ and it follows from Theorem 2.4 in [ER93] that $x_1, \ldots, x_n$ are algebraically independent, and therefore by Theorem 8.1, x has full rank. □

It follows that our constructions of rank extractors can be viewed as 'dispersers' for low-degree sources over $\mathbb{C}$. That is, they are fixed mappings that map every k-dimensional low degree source over $\mathbb{C}^n$ into almost all of $\mathbb{C}^k$.

Corollary 8.3. *Fix any integers d, k and n with $n \geq k$. Let $y : \mathbb{C}^n \to \mathbb{C}^k$ be the mapping from Theorem 1. Then, for any $x \in \mathcal{M}(\mathbb{C}^k \to \mathbb{C}^n, d)$ with full rank, $y(x(\mathbb{C}^k))$ contains all of $\mathbb{C}^k$ except for a set of measure zero.*

As far as we know, this kind of generalized dispersers were not considered before, and it will be interesting to find applications for them.

9 Discussion and Open Problems

Our paper invites further work in several directions[3].

- The extractors we give in this paper work when the field size is $d^{\Omega(k)}$. Extending our results to the case where the field size is polynomial in k is an interesting open problem. Building on the results of this paper it is enough to construct such an extractor for polynomial sources of full rank.

- An affine source may be viewed in two dual ways: as the image of an affine map, or as the kernel of one. Our extension here to low degree sources takes the first view. An interesting problem is extending the second view: extracting from low degree algebraic varieties.

- We prove an exponential upper bound of $(n+1)d^n$ on the degree of the annihilating polynomial for a set of degree d dependent polynomials in n variables. Can this bound be improved in general? Are there lower bounds? This seems to be open even over the complex numbers. An improvement on the upper bound above will yield a tighter connection between min-entropy and algebraic rank for smaller field sizes. However, it is possible that the latter can be obtained without the former.

- What is the computational complexity of testing algebraic independence? When the field size affords the equivalence to the rank of the Jacobian, there is a simple RP algorithm. Can one do it for smaller fields?

[3] A recent work of Kayal [Kay07] makes progress on several of these issues.

- What is the complexity of finding an annihilating polynomial when the polynomials are dependent? Our degree bound guarantees a PSPACE algorithm. Is there a better one, or can this problem be complete for this class?

10 Acknowledgments

We would like to thank Ran Raz and Amir Shpilka for numerous helpful conversations. We thank Jean Bourgain for bringing to our attention the results of Wooley [Woo96] and Bombieri [Bom66]. We are grateful to Salil Vadhan for his idea presented in Section 6. We thank Andrej Bogdanov and Gil Alon for helpful conversations.

References

[BIW04] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness using few independent sources. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 384–393, Washington, DC, USA, 2004. IEEE Computer Society.

[BKS+05] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating independence: new constructions of condensers, ramsey graphs, dispersers, and extractors. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 1–10, New York, NY, USA, 2005. ACM Press.

[Blu86] M. Blum. Independent unbiased coin flips from a correlated biased source: a finite state markov chain. *Combinatorica*, 6(2):97–108, 1986.

[Bom66] E. Bombieri. On exponential sums in finite fields. *American Journal of Mathematics*, 88:71–105, 1966.

[Bou07] J. Bourgain. On the construction of affine extractors. *Geometric And Functional Analysis*, 17(1):33–57, 2007.

[BRSW06] B. Barak, A. Rao, R. Shaltiel, and A. Wigderson. 2-source dispersers for sub-polynomial entropy and ramsey graphs beating the frankl-wilson construction. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 671–680, New York, NY, USA, 2006. ACM Press.

[CG88] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, April 1988. Special issue on cryptography.

[CGH+85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem of t-resilient functions (preliminary version). In *Proceedings of the 26th Annual Symposium on Foundations of Computer Science*, pages 396–407, 1985.

[CLO92] D. Cox, J. Little, and D. O'shea. *Ideals, Varieties and Algorithms.* Springer, 1992.

[ER93] R. Ehrenborg and G. Rota. Apolarity and canonical forms for homogeneous polynomials. *Europ. J. Combinatorics*, 14:157–181, 1993.

[Gan59] F. R. Gantmacher. *The Theory of Matrices*, volume 1. New York, NY, USA, 1959.

[Gol95] O. Goldreich. Three XOR-lemmas - an exposition. *Electronic Colloquium on Computational Complexity (ECCC)*, 2(056), 1995.

[GR05] A. Gabizon and R. Raz. Deterministic extractors for affine sources over large fields. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 407–418, Washington, DC, USA, 2005. IEEE Computer Society.

[GRS04] A. Gabizon, R. Raz, and R. Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 394–403, Washington, DC, USA, 2004. IEEE Computer Society.

[Har92] J. Harris. *Algebraic Geometry - A First Course.* Springer, 1992.

[Ind07] P. Indyk. Uncertainty principles, extractors, and explicit embeddings of l2 into l1. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 615–620, 2007.

[Kay07] N. Kayal. The complexity of the annihilating polynomial. *Manuscript*, 2007.

[KRVZ06] J. Kamp, A. Rao, S. Vadhan, and D. Zuckerman. Deterministic extractors for small-space sources. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 691–700, New York, NY, USA, 2006. ACM Press.

[KZ03] J. Kamp and D. Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, 2003.

[LN97] R. Lide and H. Niederreiter. *Finite fields.* Cambridge University Press, New York, NY, USA, 1997.

[L'v84] M. S. L'vov. Calculation of invariants of programs interpreted over an integrality domain. *Kibernetika*, (4):23–28, 1984.

[NZ93] N. Nisan and D. Zuckerman. More deterministic simulation in logspace. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 235–244, New York, NY, USA, 1993. ACM Press.

[Rao06] A. Rao. Extractors for a constant number of polynomially small min-entropy independent sources. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 497–506, New York, NY, USA, 2006. ACM Press.

[Rao07] Anup Rao. An exposition of bourgain's 2-source extractor. Technical Report TR07-034, ECCC, 2007.

[Raz05] R. Raz. Extractors with weak random seeds. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 11–20, New York, NY, USA, 2005. ACM Press.

[RRV99] R. Raz, O. Reingold, and S. Vadhan. Error reduction for extractors. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, page 191, Washington, DC, USA, 1999. IEEE Computer Society.

[Sch80] J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980.

[Sha94] I. R. Shafarevich. *Basic algebraic geometry.* Springer-Verlag New York, Inc., New York, NY, USA, 1994.

[Sha02] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.

[TSZ01] A. Ta-Shma and D. Zucherman. Extractor codes. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 193–199, New York, NY, USA, 2001. ACM Press.

[TV00] L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, page 32, Washington, DC, USA, 2000. IEEE Computer Society.

[Vaz87] U. Vazirani. Strong communication complexity or generating quasi-random sequences from two communicating semi-random sources. *Combinatorica*, 7:375–392, 1987.

[vN51] J. von Neumann. Various techniques used in connection with random digits. *Applied Math Series*, 12:36–38, 1951.

[Woo96] T. Wooley. A note on simultaneous congruences. *J. Number Theory*, 58:288–297, 1996.

[WZ99] A. Wigderson and D. Zuckerman. Expanders that beat the eigenvalue bound: Explicit construction and applications. *Combinatorica*, 19(1):125–138, 1999.

[Zip79] R. Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposiumon on Symbolic and Algebraic Computation*, pages 216–226. Springer-Verlag, 1979.

A Basic Notions From Algebraic Geometry

In Section 5 we use a theorem of Bombieri [Bom66] regarding character sums over curves. The very statement, let alone applicability of Bombieri's theorem requires some basic notions from algebraic geometry. In this section, we give some basic background necessary for stating the theorem and applying it as done in Section 5. The main issue in Section 5 is to show that the varieties that come up in that section are suitable for the theorem. Specifically, we need to show that these varieties are indeed curves, i.e., have dimension 1 and that their 'degree' is not too large. (All these terms will be defined formally.) For this purpose, we need some lemmas regarding the dimension and degree of intersections of varieties. Another issue is that Bombieri's theorem is stated for projective curves while we want to apply it on affine curves. For this purpose, we need some lemmas on the relations between affine and projective varieties. We note that all these issues are standard. We stress that this section is far from a full introduction to basic algebraic geometry. For a very accessible introduction we recommend [CLO92] whom most of the the definitions and notation in this section follow. Throughout this section $\mathbb{F}$ will always denote an algebraically closed field.

A.1 Affine and projective varieties

The basic objects of study in algebraic geometry are the sets of solutions to a system of polynomial equations. Such a set is called a *variety*. We now formally define affine space and affine varieties.

Definition A.1 (Affine space). *We define n-dimensional affine space over $\mathbb{F}$ as*[4]

$$\mathbb{F}^n \triangleq \{(a_1, \dots, a_n) \mid a_i \in \mathbb{F}\}$$

Definition A.2 (Affine variety). *Let $f_1, \dots, f_s$ be polynomials in $\mathbb{F}[x_1, \dots, x_n]$. We set*

$$\mathbf{V}(f_1, \dots, f_s) = \{(a_1, \dots, a_n) \in \mathbb{F}^n \mid \forall\, 1 \le i \le s\ f_i(a_1, \dots, a_n) = 0\}.$$

We call $\mathbf{V}(f_1, \dots, f_s)$ the affine variety *defined by $f_1, \dots, f_s$. A subset $V \subseteq \mathbb{F}^n$ is an* affine variety *if $V = \mathbf{V}(f_1, \dots, f_s)$ for some set of polynomials $f_1, \dots, f_s \in \mathbb{F}[x_1, \dots, x_n]$. We say that V is* reducible *if it can be written as $V = V_1 \cup V_2$ where the V_i's are affine varieties such that $V \neq V_1, V_2$. Otherwise, we say that V is* irreducible.[5]

As a simple example of an affine variety take $V = \mathbf{V}(x_1 \cdot x_2) \subseteq \mathbb{F}^2$. Note that V is reducible as it is the union of the varieties $V_1 = \mathbf{V}(x_1)$ and $V_2 = \mathbf{V}(x_2)$, i.e., the sets $\{(0, x_2) | x_2 \in \mathbb{F}\}, \{(x_1, 0) | x_1 \in \mathbb{F}\} \subseteq \mathbb{F}^2$. It can be shown that V_1 and V_2 are irreducible. Note that this is not a disjoint union as $V_1 \cap V_2 = (0, 0)$.

Though affine space and affine varieties seem to be the natural objects we want to investigate, it turns out to be very useful to work in *projective space*. Intuitively, a projective space is an affine space extended with additional 'extra points'. This intuition may not be clear from the following definition but will become clearer later on.

Definition A.3 (Projective space). *We define an equivalence relation $\sim$ over $\mathbb{F}^{n+1} \backslash \{0\}$ by setting*

$$(x_0, \dots, x_n) \sim (y_0, \dots, y_n)$$

if and only if there exists a nonzero $\lambda \in \mathbb{F}$ such that $(x_0, \dots, x_n) = (\lambda \cdot y_0, \dots, \lambda \cdot y_n)$. We define the n-dimensional projective space $\mathbb{P}^n$ over $\mathbb{F}$ to be the set of all equivalence classes of $\sim$. Thus,

$$\mathbb{P}^n = (\mathbb{F}^{n+1} - \{0\}) / \sim .$$

Each non-zero $n+1$-tuple $(x_0, \dots, x_n) \in \mathbb{F}^n$ defines a point $p \in \mathbb{P}^n$. We say that $(x_0, \dots, x_n)$ are homogenous coordinates *of p.*

[4] In most textbooks in algebraic geometry the notation $\mathbb{A}^n$ is used rather than $\mathbb{F}^n$. However, in [CLO92] which we are following, the more usual $\mathbb{F}^n$ is used.

[5] In many textbooks, the term variety always means an irreducible variety and general varieties are called *algebraic sets*.

We say that a polynomial $f \in \mathbb{F}[x_0, \ldots, x_n]$ is *homogenous* if all of its monomials have the same total degree. It is easy to see that for a homogenous polynomial f of total degree d and any nonzero $\lambda \in \mathbb{F}$

$$f(\lambda \cdot a_0, \ldots, \lambda \cdot a_n) = \lambda^d f(a_0, \ldots, a_n).$$

In particular, $f(\lambda \cdot a_0, \ldots, \lambda \cdot a_n) = 0$ if and only if $f(a_0, \ldots, a_n) = 0$. Thus, the set of 'zeros' of f is a well defined object in $\mathbb{P}^n$.

This leads to the following definition.

Definition A.4 (Projective variety). *Let $f_1, \ldots, f_s \in \mathbb{F}[x_0, \ldots, x_n]$ be homogenous polynomials. We set*

$$\mathbf{V}(f_1, \ldots, f_s) = \{(a_0, \ldots, a_n) \in \mathbb{P}^n \mid \forall\, 1 \leq i \leq s\ f_i(a_0, \ldots, a_n) = 0\}$$

A subset $V \subseteq \mathbb{P}^n$ is a projective variety *if $V = \mathbf{V}(f_1, \ldots, f_s)$ for some set of homogenous polynomials $f_1, \ldots, f_s \in \mathbb{F}[x_0, \ldots, x_n]$. We say that V is* reducible *if it can be written as $V = V_1 \cup V_2$ where the V_i's are projective varieties such that $V \neq V_1, V_2$. Otherwise, we say that V is* irreducible.

An important basic property of (affine and projective) varieties is that they decompose into irreducible varieties in a unique way. Thus, we can speak unambiguously about the irreducible components of a variety.

Proposition A.5. *-[[CLO92], Chapter 4, §6, Theorem 4 and Chapter 8, §3, Theorem 6] We say that $V = V_1 \cup \ldots \cup V_m$ is a* minimal decomposition *of V if $V_i \nsubseteq V_j$ for every $i \neq j$. Let V be an affine (projective) variety. Then V has a minimal decomposition*

$$V = V_1 \cup \ldots \cup V_m$$

where the V_i's are irreducible affine (projective) varieties. Furthermore, this minimal decomposition is unique up to the order in which $V_1, \ldots, V_m$ are written.

A.2 Varieties and ideals

An affine variety is essentially a geometric object - a set of points in the space $\mathbb{F}^n$. A fundamental idea in algebraic geometry is to relate a variety to an algebraic object. This algebraic object will be the set of all polynomials that vanish on the variety. It is easy to see that this set of polynomials forms an ideal in the ring $\mathbb{F}[x_1, \ldots, x_n]$. First we recall some basic facts and notation regarding ideals in $\mathbb{F}[x_1, \ldots, x_n]$. For $f_1, \ldots, f_s \in \mathbb{F}[x_1, \ldots, x_n]$ we denote by $< f_1, \ldots, f_s >$ the ideal generated by $f_1, \ldots, f_s$. That is,

$$< f_1, \ldots, f_s > \triangleq \{\sum_{i=1}^{s} g_i \cdot f_i \mid \forall\, 1 \leq i \leq s\ g_i \in \mathbb{F}[x_1, \ldots, x_n]\}.$$

By the Hilbert Basis Theorem (see [CLO92], Chapter 2, §5) every ideal $I \subset \mathbb{F}[x_1, \ldots, x_n]$ is *finitely generated*, i.e., $I =< f_1, \ldots, f_s >$ for some $f_1, \ldots, f_s \in \mathbb{F}[x_1, \ldots, x_n]$. For an ideal $I =< f_1, \ldots, f_s >$, it is easy to see that a point $(a_1, \ldots, a_n) \in \mathbb{F}^n$ is a zero of every $f \in I$ if and only if it is a zero of $f_1, \ldots, f_s$.

Definition A.6 (Affine varieties and ideals). *For an affine variety $V \subseteq \mathbb{F}^n$ we define $\mathbf{I}(V)$ to be the ideal of all polynomials f such that $f(a_1, \ldots, a_n) = 0$ for every $(a_1, \ldots, a_n) \in V$. For* *an ideal $I =< f_1, \dots, f_s >\subseteq \mathbb{F}[x_1, \dots, x_n]$ we define $\mathbf{V}(I) \subseteq \mathbb{F}^n$ to be the affine variety $\mathbf{V}(I) = \{(a_1, \dots, a_n) \mid f(a_1, \dots, a_n) = 0 \ , \forall f \in I\} = \mathbf{V}(f_1, \dots, f_s)$.*

Before making the corresponding definitions for projective varieties we will need some terminology. We remarked above that it makes sense to ask whether a homogenous polynomial $f \in \mathbb{F}[x_0, \dots, x_n]$ vanishes at a point $p \in \mathbb{P}^n$. For a non-homogenous polynomial $f \in \mathbb{F}[x_0, \dots, x_n]$ we say that $f(p) = 0$ for $p \in \mathbb{P}^n$ if $f(a_0, \dots, a_n) = 0$ for all representatives $(a_0, \dots, a_n)$ of p.

We say that an ideal $I \subseteq \mathbb{F}[x_0, \dots, x_n]$ is *homogenous* if it is generated by a set of homogenous polynomials, i.e., $I =< f_1, \dots, f_s >$ where $f_1, \dots, f_s$ are homogenous. We can now give the following definitions.

Definition A.7 (Projective varieties and homogenous ideals). *For a projective variety $X \subseteq \mathbb{P}^n$ we define $\mathbf{I}(X)$ to be the ideal of all polynomials f with $f(p) = 0$ for every $p \in X$. It can be shown that $\mathbf{I}(X)$ is always a homogenous ideal.*
For a homogenous ideal $I \subseteq \mathbb{F}[x_0, \dots, x_n]$ we define $\mathbf{V}(I) \subseteq \mathbb{P}^n$ to be the projective variety of all points $p \in \mathbb{P}^n$ that are zeros of all polynomials $f \in I$. If $I =< f_1, \dots, f_s >$ for homogenous polynomials $f_1, \dots, f_s$ then it can be shown that $\mathbf{V}(I) = \mathbf{V}(f_1, \dots, f_s)$.

One reason the correspondence between ideals and varieties is useful is that operations on ideals have simple corollaries in terms of the corresponding varieties. We need the following fact about intersections of ideals.

Proposition A.8 ([CLO92], Chapter 4, §3, Theorem 15 and Chapter 8, §3, Exercise 7). *Let I_1, I_2 be ideals in $\mathbb{F}[x_1, \dots, x_n]$ or homogenous ideals in $\mathbb{F}[x_0, \dots, x_n]$. Then*

$$\mathbf{V}(I_1 \cap I_2) = \mathbf{V}(I_1) \cup \mathbf{V}(I_2).$$

A.3 The dimension and degree of a variety

There are several equivalent definitions of the dimension and degree of a variety (degree is defined only for projective varieties). Here we define dimension and degree in terms of the Hilbert polynomial of a variety. First we need to define the Hilbert function and Hilbert polynomial of an ideal.

We say that an ideal I is a *monomial ideal* if it is generated by a set of monomials.[6] For example $I =< x_1, x_2^2 >$ is a monomial ideal. We first define the Hilbert function for monomial ideals.

Definition A.9 (Hilbert function of a monomial ideal). *Let I be a monomial ideal in $\mathbb{F}[x_1, \dots, x_n]$. The* affine Hilbert function *of I denoted ${}^aHF_I(s)$, is a function on non-negative integers defined by ${}^aHF_I(s)$ = number of monic monomials in $\mathbb{F}[x_1, \dots, x_n]$ of degree at most s not contained in I.*
Similarly, let I be a homogenous monomial ideal in $\mathbb{F}[x_0, \dots, x_n]$. The Hilbert function *of I denoted $HF_I(s)$, is a function on non-negative integers defined by*
$HF_I(s)$ = number of monic monomials in $\mathbb{F}[x_0, \dots, x_n]$ of degree exactly *s not contained in I.*

Roughly speaking, for a monomial ideal I the monomials not in I are a basis for the space of polynomials that are 'different modulo I'. Thus, ${}^aHF_I(s)$ is the dimension of the space of such

[6] By Dickson's Lemma ([CLO92], Chapter 2, §4 Theorem 5) if I is a monomial ideal it can always be generated by a finite set of monomials.

polynomials of degree at most s. This is the idea behind the definition of the Hilbert function for general ideals. First we need some notation. For a subset of polynomials $V \subseteq \mathbb{F}[x_1, \ldots, x_n]$ and a non-negative integer s, we denote by $V_{\leq s} \subseteq \mathbb{F}[x_1, \ldots, x_n]$ the set of polynomials in V of (total) degree at most s. For example, $\mathbb{F}[x_1, \ldots, x_n]_{\leq s}$ is the set of all polynomials of degree at most s. Similarly, for a subset $V \subseteq \mathbb{F}[x_0, \ldots, x_n]$ we denote by $V_s \subseteq \mathbb{F}[x_0, \ldots, x_n]$ the set of all polynomials in V of degree *exactly* s. Note that if $V \subseteq \mathbb{F}[x_1, \ldots, x_n]$ is a linear subspace, then so are $V_{\leq s}$ and V_s. In particular if $I \subseteq \mathbb{F}[x_1, \ldots, x_n]$ is an ideal, then it is also a linear subspace and so is $V_{\leq s}$. We recall a basic notion for linear algebra: For subspaces $W \subseteq V \subseteq \mathbb{F}[x_1, \ldots, x_n]$ we denote by V/W the *quotient space* of equivalence classes of V over W. That is, we define an equivalence relation $\sim$ over V by $v \sim v' \leftrightarrow v - v' \in W$ and let V/W be the space of these equivalence classes. We can now make the following definition.

Definition A.10 (Hilbert function of a general ideal). *Let I be an ideal in $\mathbb{F}[x_1, \ldots, x_n]$. The* affine Hilbert function *of I, denoted ${}^aHF_I(s)$, is defined as ${}^aHF_I(s) \triangleq dim\,(\mathbb{F}[x_1, \ldots, x_n]_{\leq s}/I_{\leq s})$.*
Let I be a homogenous ideal in $\mathbb{F}[x_0, \ldots, x_n]$ the Hilbert function *of I, denoted $HF_I(s)$ is defined as $HF_I(s) \triangleq dim\,(\mathbb{F}[x_0, \ldots, x_n]_s/I_s)$.*

It can be shown that for large enough input s, the Hilbert Function coincides with a polynomial.

Theorem A.11 (See [CLO92] Chapter 9, §3). .

1. *Let I be an ideal in $\mathbb{F}[x_1, \ldots, x_n]$. There exists a polynomial ${}^aHP_I(s)$ such that for large enough s, ${}^aHP_I(s) = {}^aHF_I(s)$. We call ${}^aHP_I(s)$ the* affine Hilbert polynomial *of I.*

2. *Let I be a homogenous ideal in $\mathbb{F}[x_0, \ldots, x_n]$. There exists a polynomial $HP_I(s)$ such that for large enough s, $HP_I(s) = HF_I(s)$. We call $HP_I(s)$ the* Hilbert polynomial *of I.*

Let $V \subseteq \mathbb{F}^n$ be an affine variety with $I = \mathbf{I}(V)$. Let's try to see why it could make sense to define the dimension of a variety in terms of the affine Hilbert polynomial of I. Since I is exactly the set of polynomials that vanish on V, polynomials $f, g \in \mathbb{F}[x_1, \ldots, x_n]$ are identical on V if and only if $f - g \in I$. It follows that $\mathbb{F}[x_1, \ldots, x_n]/I$ is exactly the space of polynomial functions from V to $\mathbb{F}$. Now recall that for a linear subspace $A \subseteq \mathbb{F}^n$, the dimension of A can be defined as the dimension of the space of linear functions from A to $\mathbb{F}$. Similarly, we could try to define the dimension of V as the dimension of the space of *polynomial* functions from V to $\mathbb{F}$, i.e., the dimension of $\mathbb{F}[x_1, \ldots, x_n]/I$. However, since the polynomials in this space have unbounded degree, $\mathbb{F}[x_1, \ldots, x_n]/I$ has infinite dimension. Instead, we can take an 'asymptotic' approach and define the dimension of V by how fast this space grows as we increase the degree of the polynomials. More accurately, we can define $dim(V)$ by how fast ${}^aHP_I(s) = dim(\mathbb{F}[x_1, \ldots, x_n]_{\leq s}/I_{\leq s})$ grows as s increases. This corresponds to the degree of ${}^aHP_I(s)$.

Definition A.12 (Dimension of a variety). *Let $V \subseteq \mathbb{F}^n$ be an affine variety and let $I = \mathbf{I}(V)$. The* dimension *of V denoted $dim(V)$, is defined to be the degree of ${}^aHP_I(s)$.*
Let $V \subseteq \mathbb{P}^n$ be a projective variety and let $I = \mathbf{I}(V)$. The dimension *of V is defined to be the degree of $HP_I(s)$.*

To gain intuition on the above definition, it is helpful to see how it coincides with the definition of dimension for a linear subspace. Take for example, the subspace $V \subseteq \mathbb{F}^n$ defined by the constraints

$\{x_1 = 0,\ x_2 = 0\}$. Then $I \triangleq \mathbf{I}(V) =< x_1, x_2 >$ and the monomials *not* in I are exactly the monomials $x_3^{a_3} \cdots x_n^{a_n}$ where $a_3, \ldots, a_n$ are non-negative integers. In particular, the number of such monomials of degree at most s is $\binom{n-2+s}{n-2}$, which is a degree $n-2$ polynomial in s. Therefore, since I is a monomial ideal by the definition above $dim(V) = deg(HP_I(s)) = n-2$.

The following property of the dimension of a variety will be very useful for us later on.

Proposition A.13 ([CLO92], Chapter 9,§4 Corollary 9). *Let V be an affine or projective variety. Then the dimension of V is the maximum of the dimensions of its irreducible components.*

We now define the *degree* of a projective variety (degree is not defined for affine varieties).

Definition A.14 (Degree of a variety). *The* degree *of V denoted $deg(V)$, is defined to be the leading coefficient of $HP_I(s)$ multiplied by $dim(V)!$.*

Though not immediate from the definition, it can be shown that the degree is always a non-negative integer. To gain intuition on the above definition, let us see how it coincides with the definition of degree for a univariate polynomial. For simplicity of the example we'll assume degree is defined for an affine variety V in a similar way to projective varieties. That is, $deg(V)$ is the leading coefficient of the affine Hilbert polynomial of $\mathbf{I}(V)$ times $dim(V)!$. Let $I \subseteq \mathbb{F}[x_1]$ be the ideal $< x_1^3 - 1 >$. It can be shown that $I = \mathbf{I}(V)$ where $V = \mathbf{V}(x_1^3 - 1) \in \mathbb{F}$, i.e., V is simply the roots of $x_1^3 - 1$ and $|V| = 3$ (since $\mathbb{F}$ is algebraically closed). Furthermore, it can be seen that $\{1, x_1, x_1^2\}$ is a basis for $k[x_1]/I$. Hence, $HP_I(s)$ is simply the constant 3 and therefore $dim(V) = deg(HP_I(s)) = 0$ and $deg(V) = 3 \cdot 0! = 3$. Thus $deg(V)$ bounds the size of V. It can be shown that $deg(V)$ always bounds $|V|$ when V is a projective variety of finite size.

A.4 The projective closure of an affine variety

We call an affine (projective) variety of dimension 1 an affine (projective) curve. As mentioned above, in Section 5 we use a theorem of Bombieri [Bom66] for affine curves while in [Bom66] the theorem is stated for projective curves. The transition between the cases, presented in subsection A.7, is completely standard. For this purpose, the following definitions enable us to relate an affine variety with its 'corresponding' projective variety. First we need the following definitions.

Definition A.15 (Homogenization). .

- *For a polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ of degree d, we define the homogenized version $f^h \in \mathbb{F}[x_0, \ldots, x_n]$ by*
$$f^h(x_0, x_1, \ldots, x_n) = x_0^d \cdot f(x_1/x_0, \ldots, x_n/x_0).$$
- *Similarly, for an ideal $I =< f_1, \ldots, f_s >$ we define the ideal $I^h =< f^h | f \in I >$. Note that I^h is always homogenous. In particular, it is easy to see that $I^h =< f_1^h, \ldots, f_s^h >$.*

We can now define the projective closure of an affine variety.

Definition A.16 (Projective closure). *Let $V \subseteq \mathbb{F}^n$ be an affine variety with ideal $I = \mathbf{I}(V)$. We define the* projective closure $\overline{V} \subseteq \mathbb{P}^n$ *to be the projective variety $\mathbf{V}(I^h)$.*
Let $U_0 \subseteq \mathbb{P}^n$ be defined as $U_0 = \{(a_0, a_1, \ldots, a_n) \in \mathbb{P}^n | a_0 = 1\}$. Note that U_0 can be identified with

$\mathbb{F}^n$. Thus, we can think of an affine variety $V \subseteq \mathbb{F}^n$ as being contained in U_0. For a projective variety $V \subseteq \mathbb{P}^n$, we denote $V^a \triangleq V \cap U_0$. Intuitively, this is "the affine part of V".

The following propositions, show various connections between an affine variety and its projective closure.

Proposition A.17 ([CLO92]-Chapter 8, §4, Proposition 7 and Exercise 9). *Let $V \subseteq \mathbb{F}^n$ be an affine variety. Then*

1. $\overline{V} \cap U_0 = V$.
2. *V is irreducible if and only if $\overline{V}$ is irreducible.*

Proposition A.18 ([CLO92]-Chapter 9, §3, Theorem 12). *Let $V \subseteq \mathbb{F}^n$ be an affine variety. Then*

$$dim(V) = dim(\overline{V}).$$

Proposition A.19 ([CLO92]-Chapter 8, §4, Theorem 8). *Let $f_1, \ldots, f_r \in \mathbb{F}[x_1, \ldots, x_n]$ be polynomials such that $V = \mathbf{V}(f_1, \ldots, f_r) \subseteq \mathbb{F}^n$ is non-empty. Then*

$$\overline{V} = \mathbf{V}(f_1^h, \ldots, f_r^h).$$

Claim A.20. *Let $V_1, \ldots, V_r \subseteq \mathbb{F}^n$ be affine varieties. Then $\overline{V_1 \cup \ldots \cup V_r} = \overline{V}_1 \cup \ldots \cup \overline{V}_r$.*

Proof. We prove the claim for $r = 2$. The statement for general r follows by induction.

Let I_1, I_2 be the ideals $\mathbf{I}(V_1), \mathbf{I}(V_2)$ respectively. It can be shown that $\mathbf{V}(I_1^h \cap I_2^h) = \mathbf{V}((I_1 \cap I_2)^h)$. We have

$$\overline{V_1 \cup V_2} = \mathbf{V}((I_1 \cap I_2)^h) = \mathbf{V}(I_1^h \cap I_2^h) = \mathbf{V}(I_1^h) \cup \mathbf{V}(I_2^h) = \overline{V}_1 \cup \overline{V}_2,$$

where we used Proposition A.8 in the first and second to last equality.

□

Corollary A.21. *Let $V \subseteq \mathbb{F}^n$ be an affine variety with irreducible components $V_1, \ldots, V_r$. Then, the irreducible components of $\overline{V} \subseteq \mathbb{P}^n$ are $\overline{V}_1, \ldots, \overline{V}_r$.*

Proof. Follows from Proposition A.17 and Claim A.20. □

Claim A.22. *Let $V \subseteq \mathbb{F}^n$ be an affine variety. If $f \in \mathbb{F}[x_1, \ldots, x_n]$ does not vanish identically on V then f^h does not vanish identically on $\overline{V} \subseteq \mathbb{P}^n$.*

Proof. For any $a \in \mathbb{F}^n$ $f(a) = f^h(1, a)$. Therefore, if $f(a) \neq 0$ for $a \in V$, then $f^h(1, a) \neq 0$ where $(1, a) \in \overline{V}$ by Proposition A.17. □

A.5 The dimension of intersections of hypersurfaces

We say that an affine (projective) variety V is *hypersurface* if $V = \mathbf{V}(f)$ for a (homogenous) polynomial f. In this subsection we state and prove standard results regarding the dimension of intersections of hypersurfaces. The following definition will be important.

Definition A.23. *We say that an affine or projective variety V has* pure dimension *if all its irreducible components have the same dimension.*

We need the following propositions about the intersection of a hypersurface with a variety.

Proposition A.24 ([CLO92] Chapter 9, §4, Proposition 7). *Let $V \subseteq \mathbb{P}^n$ be a projective variety with $dim(V) \geq 1$. Then for any non-constant homogenous polynomial $f \in \mathbb{F}[x_0, \ldots, x_n]$, $V \cap \mathbf{V}(f) \neq \emptyset$.*

Proposition A.25 ([Sha94], Chapter I, §6, Corollary 1 of Theorem 5). *Let $V \subseteq \mathbb{P}^n$ be an irreducible projective variety. Let $f \in \mathbb{F}[x_0, \ldots, x_n]$ be a homogenous polynomial that does not vanish identically on V and denote $H = \mathbf{V}(f)$. If $V \cap H \neq \emptyset$, then $V \cap H$ has pure dimension $dim(V) - 1$.*

Claim A.26. *Let $V \subseteq \mathbb{P}^n$ be a projective variety of pure dimension $dim(V) \geq 1$. Let $f \in \mathbb{F}[x_0, \ldots, x_n]$ be a non-constant homogenous polynomial and let $H = \mathbf{V}(f) \subseteq \mathbb{P}^n$. Assume that f does not vanish identically on any of the irreducible components of V. Then $V \cap H$ has pure dimension $dim(V) - 1$.*

Proof. Let $V = Z_1 \cup \ldots \cup Z_k$ be the decomposition of V into irreducible components. Fix any $j \in [k]$. By Proposition A.24, $Z_j \cap H$ is non-empty, and since f does not vanish on Z_j, by Proposition A.25 all irreducible components of $Z_j \cap H$ have dimension $dim(V) - 1$. To conclude, note that the union of the irreducible components of $Z_j \cap H$ over all $j \in [k]$ is $V \cap H$ and therefore the irreducible components of $V \cap H$ are just a subset of these components (excluding a component that is contained in another). Hence, all irreducible components of $V \cap H$ have dimension $dim(V) - 1$ and the claim follows. □

As a special case we get the following.

Corollary A.27. *Let $f \in \mathbb{F}[x_0, \ldots, x_n]$ be a non-constant homogenous polynomial. Then the hypersurface $H = \mathbf{V}(f) \subseteq \mathbb{P}^n$ has pure dimension $n - 1$.*

Proof. $\mathbb{P}^n$ can be shown to be irreducible and in particular has pure dimension. Thus, using Claim A.26 with $V = \mathbb{P}^n$ we get the desired result. □

We can now state and prove the main lemma we use regarding the dimension of intersections of hypersurfaces.

Lemma A.28. *Let $0 < r < n$ be integers and let $f_1, \ldots, f_r \in \mathbb{F}[x_0, \ldots, x_n]$ be non-constant homogenous polynomials. For each $i \in [r]$, let $H_i = \mathbf{V}(f_i) \subseteq \mathbb{P}^n$ and let $V_i = \mathbf{V}(f_1, \ldots, f_i) = H_1 \cap \ldots \cap H_i$. Then*

1. *All irreducible components of the projective variety V_r have dimension at least $n - r$.*

2. *Suppose furthermore that for each $2 \leq i \leq r$, f_i does not vanish identically on any of the irreducible components of V_{i-1}. Then V_r is a projective variety of pure dimension $n - r$.*

Proof. We prove the first item by induction on r. For $r = 1$ this follows from Corollary A.27. Assume the claim for $r-1$. Let $V_{r-1} = Z_1 \cup \ldots \cup Z_k$ be the decomposition of V_{r-1} into irreducible components. Fix any $j \in [k]$. Similarly to the proof of Claim A.26, we will show that all the irreducible components of $Z_j \cap H_r$ have dimension at least $n - r$ and since the irreducible components of V_r are a subset of these, the claim follows. From the induction hypothesis we have $dim(Z_j) \geq n-(r-1)$. If f_r vanishes on Z_j then $Z_j \cap H_r = Z_j$ (which is the only irreducible component) and we are done. Otherwise, by Claim A.26 all components of $Z_j \cap H_r$ have dimension at least $n - r$.
We now prove the second item by induction on r. For $r = 1$ this is exactly Corollary A.27. Assume the claim for $r - 1$. Then by the induction hypothesis, V_{r-1} has pure dimension $n-r+1$. Therefore, by Claim A.26 $V_r = V_{r-1} \cap H_r$ has pure dimension $n - r$. □

We also need the corresponding lemma in affine space.

Lemma A.29. *Let $0 < r < n$ be integers and let $f_1, \ldots, f_r \in \mathbb{F}[x_1, \ldots, x_n]$ be non-constant polynomials. For each $i \in [r]$, let $H_i = \mathbf{V}(f_i) \subseteq \mathbb{F}^n$ and let $V_i = \mathbf{V}(f_1, \ldots, f_i) = H_1 \cap \ldots \cap H_i$. Suppose that for each $2 \leq i \leq r$, f_i does not vanish identically on any of the irreducible components of the affine variety V_{i-1}. Then, if V_r is non-empty it is an affine variety of pure-dimension $n - r$.*

Proof. For $1 \leq i \leq r$, let $X_i = \mathbf{V}(f_1^h, \ldots, f_i^h)$. By Proposition A.19, for every $1 \leq i \leq r$ $X_i = \overline{V}_i$. Therefore, by Corollary A.21 the irreducible components of X_{i-1} are simply the projective closures of the irreducible components of V_{i-1}. By Claim A.22 it follows that f_i^h does not vanish identically on any of the irreducible components of X_{i-1}. Hence, we can use Lemma A.28 and X_r is a projective variety of pure dimension $n - r$ and since $X_r = \overline{V}_r$, using Proposition A.18 V_r is an affine variety of pure dimension $n - r$. □

A.6 The degree of intersections of hypersurfaces

We now discuss degree. The main result we prove is the following corollary of Bezout's Theorem.

Lemma A.30. *Let $f_1, \ldots, f_r \in \mathbb{F}[x_0, \ldots, x_n]$ be non-constant homogenous polynomials of degrees $d_1, \ldots, d_r$ respectively, and let $D = d_1 \cdots d_r$. Let $X = \mathbf{V}(f_1, \ldots, f_r) \subseteq \mathbb{P}^n$. Assume that $dim(X) = n-r$. Then*

1. *$deg(X) \leq D$.*
2. *The number of irreducible components of X is at most D.*

Using this Lemma, we immediately get a bound on the number of irreducible components of an affine variety.

Lemma A.31. *Let $f_1, \ldots, f_r \in \mathbb{F}[x_1, \ldots, x_n]$ be non-constant polynomials of degrees $d_1, \ldots, d_r$, respectively, and let $D = d_1 \cdots d_r$. Let $V = \mathbf{V}(f_1, \ldots, f_r) \subseteq \mathbb{F}^n$. Assume that V is non-empty and $dim(V) = n - r$. Then the number of irreducible components of V is at most D.*

Proof. Let $X = \overline{V}$. By Proposition A.19, $X = \mathbf{V}(f_1^h, \ldots, f_r^h)$. Therefore, by Lemma A.30, X has at most D irreducible components and by Corollary A.21 V has at most D irreducible components. □

The following proposition states that a degree of a hypersurface is at most the degree of any polynomial defining it.

Proposition A.32 ([CLO92], Chapter 9, §4, Exercise 12). *Let f be a non-constant homogenous polynomial. Let $H = \mathbf{V}(f_1)$. Then $deg(H) \leq deg(f)$.*

We will need the following definitions taken from [Har92].

Definition A.33. *Let $X, Y \subseteq \mathbb{P}^n$ be projective varieties. We say that X and Y* intersect properly *if*

$$dim(X \cap Y) = dim(X) + dim(Y) - n.$$

We quote (a corollary of) Bezout's Theorem.

Theorem A.34 (Bezout-[Har92] Chapter 18, Theorem 18.4 and Corollary 18.5). *Let $X, Y \subseteq \mathbb{P}^n$ be projective varieties of pure dimension intersecting properly. Then*

1. *$deg(X \cap Y) \leq deg(X) \cdot deg(Y)$.*
2. *The number of irreducible components of $X \cap Y$ is at most $deg(X) \cdot deg(Y)$.*

Claim A.35. *Let $X = \mathbf{V}(f_1, \ldots, f_r) \subseteq \mathbb{P}^n$ where $f_1, \ldots, f_r \in \mathbb{F}[x_0, \ldots, x_n]$ are non-constant homogenous polynomials. Assume that $dim(X) = n - r$. For $i = 1, \ldots, r$ let $H_i = \mathbf{V}(f_i)$ and $X_i = \mathbf{V}(f_1, \ldots, f_i) = H_1 \cap \ldots \cap H_i$. Then for all $i \in [r]$, X_i has pure dimension $n - i$.*

Proof. By Lemma A.28, all irreducible components of $\mathbf{V}(f_1, \ldots, f_i)$ have dimension at least $n - i$. Thus, it is enough to prove that $\mathbf{V}(f_1, \ldots, f_i)$ has (not necessarily pure) dimension $n - i$. We prove this by backwards induction on i. For $i = r$ it is already given that $dim(X) = dim(X_r) = n - r$. Assume the claim for $i + 1$ and assume for contradiction that $dim(X_i) \neq n - i$. Using Lemma A.28 it follows that $dim(X_i) > n - i$. Therefore, by Claim A.26 $dim(X_{i+1}) = dim(X_i \cap \mathbf{V}(f_{i+1})) > n - (i+1)$ and this contradicts the induction hypothesis. □

We can now prove Lemma A.30.

Proof. (of Lemma A.30). We prove the claim by induction on r. For $r = 1$, it follows from Proposition A.32 that $deg(X) \leq deg(f_1) = d_1$. Assume the claim for $r - 1$. For $i = 1, \ldots, r$ denote $H_i = \mathbf{V}(f_i)$. Given $H_1, \ldots, H_r$, denote $X_{r-1} = H_1 \cap \ldots \cap H_{r-1}$. We know from the induction hypothesis that

$$deg(X_{r-1}) \leq d_1 \cdots d_{r-1}.$$

From Claim A.35, X_{r-1} has pure dimension $n - (r - 1)$ and it follows that X_{r-1} and H_r intersect properly. Therefore, we can use Theorem A.34 and get

$$\deg(X) = deg(X_{r-1} \cap H_r) \leq deg(X_{r-1}) \cdot \deg(H_r) \leq d_1 \cdots d_r = D.$$

Similarly, from Theorem A.34 we get that the number of irreducible components of X is at most $deg(X_{r-1}) \cdot deg(H_r) \leq D$. □

A.7 Bombieri's theorem

We quote an estimate of Bombieri [Bom66] for character sums over projective curves and show that the estimate can be used also for affine curves. (Recall that curve is a variety of dimension 1.) First we introduce some notation. Let $X \subseteq \mathbb{P}^n$ be a projective curve of degree D. Let $\mathbb{F}$ denote the algebraic closure of $\mathbb{F}_p$ for some prime p. Let $R \in \mathbb{F}_p(x_0, \ldots, x_n)$ be a homogenous rational function whose numerator and denumerator both have degree d. Then, for any $x \in \mathbb{F}^{n+1}$ and non-zero $\lambda \in \mathbb{F}$ we have

$$R(\lambda \cdot x) = \frac{p(\lambda \cdot x)}{q(\lambda \cdot x)} = \frac{\lambda^d p(x)}{\lambda^d q(x)}$$

$$= \frac{p(x)}{q(x)} = R(x).$$

Therefore R is a well defined function on points of $\mathbb{P}^n$ that are not poles of R, i.e., points $x \in \mathbb{P}^n$ such that $q(x) \neq 0$. We define

$$S_m(R, X) \triangleq \sum_{x \in X_m, q(x) \neq 0} e_p(\sigma R(x))$$

where X_m is the set of points of X with coordinates in $\mathbb{F}_{p^m}$, σ denotes the trace[7] from $\mathbb{F}_{p^m}$ to $\mathbb{F}_p$ and $e_p(x)$ is the function $e^{2\pi i x/p}$. Note that we sum only over non-poles of R.

Theorem A.36 (Theorem 6 in [Bom66]). *Let R and X be as above. Let $\Gamma_1, \ldots, \Gamma_L$ be the irreducible components of X. Assume R is non-constant on Γ_i for $i = 1, \ldots, L$. If $d \cdot D < p$ then*

$$|S_m(R, X)| \leq 4dD^2 \cdot p^{m/2}.$$

For an affine curve $C \subseteq \mathbb{F}^n$, and a polynomial $g \in \mathbb{F}_p[x_1, \ldots, x_n]$ we define

$$S_m(g, C) \triangleq \sum_{(a_1, \ldots, a_m) \in C_m} e_p(\sigma g(a_1, \ldots, a_m))$$

where C_m denotes the set of points of C with coordinates in $\mathbb{F}_{p^m}$. We also denote $S(g, C) \triangleq S_1(g, C)$. We can now state and prove a version of Theorem A.36 for affine curves.

Theorem A.37. *Let $V \subseteq \mathbb{F}^n$ be an affine curve such that $V = \mathbf{V}(f_1, \ldots, f_{n-1})$ for polynomials $f_i \in \mathbb{F}[x_1, \ldots, x_n]$. Let $D = deg(f_1) \cdots deg(f_{n-1})$. Let $V_1, \ldots, V_L$ be the irreducible components of V. Let $g \in \mathbb{F}_p[x_1, \ldots, x_n]$ be a polynomial of degree d that is non-constant on some V_i. Let C be the union of the irreducible components V_i such that g is non-constant on V_i. Assume that $d \cdot D < p$. We have*

$$S_m(g, C) \leq 4dD^2 \cdot p^{m/2}.$$

In particular,

$$S(g, C) \leq 4dD^2 \cdot p^{1/2}.$$

[7] See [LN97] for a definition of the trace function. For the case $m = 1$, which is the only one we will use, the trace is simply the identity function.

Proof. We identify g with a homogenous rational function R defined as

$$R(x_0, x) = \frac{g^h(x_0, x)}{x_0^d}$$

Note that for every $a \in \mathbb{F}^n$ $R(1, a) = g(a)$.

Denote $X = \overline{C}$.

Claim A.38.

$$S_m(g, C) = S_m(R, X).$$

Proof. Using Proposition A.17 X consists precisely of the points $(1, a)$ where $a \in C$ and, possibly, some 'points at infinity', i.e., points of the form $(0, a)$ for $a \in \mathbb{F}^n$. Since R has poles on all points of the form $(0, a)$ and $R(1, a) = g(a)$ for all $x \in \mathbb{F}^n$, we get that summing R over all non-poles in X is exactly the same as summing g over all of C. In particular, summing R over all non-poles in X_m is exactly the same as summing g over all of C_m. That is,

$$S_m(g, C) = S_m(R, X).$$

□

We now want to bound $S_m(R, X)$ using Theorem A.36. Note that both the numerator and denumerator of R are homogenous of degree exactly d so R is suitable for the theorem. We need to show that X is a projective variety of dimension 1 such that R is non-constant on any of its irreducible components: Recall that the irreducible components of C are simply a subset of $V_1, \ldots, V_L$. Assume without loss of generality, that $C = V_1 \cup \ldots \cup V_r$. Using Corollary A.21, it is clear that if g is non-constant on the irreducible components $V_1, \ldots, V_r$ of C, then R is non-constant on the irreducible components $\overline{V}_1, \ldots, \overline{V}_r$ of X. By Proposition A.18 and Corollary A.21 $dim(\overline{V}) = 1$ and $\overline{V}_1, \ldots, \overline{V}_L$ are the irreducible components of $\overline{V}$. By Proposition A.19, $\overline{V} = \mathbf{V}(f_1^h, \ldots, f_{n-1}^h)$ and therefore by Claim A.35 for every i, $\overline{V}_i$ has dimension 1. It follows that $X = \overline{V}_1 \cup \ldots \cup \overline{V}_r$ has dimension 1.
Finally, we need to bound the degree of X. By Lemma A.30 $deg(\overline{V}) \leq D$. Since the degree of a projective variety is the sum of degrees of its irreducible components (see [Har92], Chapter 18) then $deg(X) \leq D$.

Therefore, we can use Theorem A.36. We get

$$|S_m(g, C)| = |S_m(R, X)| \leq 4dD^2 \cdot p^{m/2}.$$

□