Extractors for Images of Varieties

Zeyu Guo
zguotcs@gmail.com
Ohio State University
Columbus, Ohio, USA

Ben Lee Volk
benleevolk@gmail.com
Reichman University
Herzliya, Israel

Akhil Jalan
akhiljalan@utexas.edu
University of Texas at Austin
Austin, Texas, USA

David Zuckerman
diz@utexas.edu
University of Texas at Austin
Austin, Texas, USA

ABSTRACT

We construct explicit deterministic extractors for *polynomial images of varieties*, that is, distributions sampled by applying a low-degree polynomial map $f : \mathbb{F}_q^r \to \mathbb{F}_q^n$ to an element sampled uniformly at random from a k-dimensional variety $V \subseteq \mathbb{F}_q^r$. This class of sources generalizes both *polynomial sources*, studied by Dvir, Gabizon and Wigderson (FOCS 2007, Comput. Complex. 2009), and *variety sources*, studied by Dvir (CCC 2009, Comput. Complex. 2012).

Assuming certain natural non-degeneracy conditions on the map f and the variety V, which in particular ensure that the source has enough min-entropy, we extract almost all the min-entropy of the distribution. Unlike the Dvir–Gabizon–Wigderson and Dvir results, our construction works over large enough finite fields of arbitrary characteristic. One key part of our construction is an improved deterministic rank extractor for varieties. As a by-product, we obtain explicit Noether normalization lemmas for affine varieties and affine algebras.

Additionally, we generalize a construction of affine extractors with exponentially small error due to Bourgain, Dvir and Leeman (Comput. Complex. 2016) by extending it to all finite prime fields of quasipolynomial size.

CCS CONCEPTS

• **Theory of computation → Pseudorandomness and derandomization**.

KEYWORDS

Pseudorandomness, Extractors, Varieties, Polynomial Maps

ACM Reference Format:
Zeyu Guo, Ben Lee Volk, Akhil Jalan, and David Zuckerman. 2023. Extractors for Images of Varieties. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC '23), June 20–23, 2023, Orlando, FL, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3564246.3585109

STOC '23, June 20–23, 2023, Orlando, FL, USA

ACM ISBN 978-1-4503-9913-5/23/06…$15.00
https://doi.org/10.1145/3564246.3585109

1 INTRODUCTION

Randomness is a powerful resource in computing. There are many useful randomized algorithms, and randomness is provably necessary in cryptography and distributed computing. Naturally, these uses of randomness assume access to uniformly random bits. However, it can be expensive or impossible to obtain such high-quality randomness. A randomness extractor converts low-quality randomness into high-quality randomness.

Low-quality random sources can arise in several ways. First, natural sources of randomness may be defective. Second, in cryptography, if an adversary gains information about a string, then conditioned on this information, the string is weakly random. Third, in constructing pseudorandom generators, a similar situation arises when we condition on the state of the computation. Besides the computer science motivation, randomness extraction questions are natural mathematically.

We model a weak source as a class $\mathcal{D}$ of distributions over a finite set Ω. A randomness extractor for $\mathcal{D}$ is a deterministic function that extracts randomness from any distribution in $\mathcal{D}$.

Definition 1.1. *An extractor for a class $\mathcal{D}$ of distributions with error ε, or an ε-extractor, is a function* $\mathsf{Ext} : \Omega \to B$ *such that for any $D \in \mathcal{D}$, the distribution* $\mathsf{Ext}(D)$ *is ε-close, in statistical distance, to the uniform distribution over B.*

Typically the codomain B will be $\{0,1\}^m$.

The most general class of distributions is the set of distributions with high min-entropy, i.e., distributions that do not place much probability on any string. However, it is not hard to show that it is impossible to extract from such sources. It is possible to extract using an auxiliary seed, and there are many applications of such seeded extractors (see [32] for a survey). It is also possible to extract from two independent general weak sources (e.g., [8]). However, if we want to avoid adding a seed and only have one source, we must restrict the class of distributions further.

Various models of weak sources have been studied. It is not hard to show that if there are not too many distributions in the class, then most functions are extractors with excellent parameters. Of course, we really want efficiently-computable extractors.

Models of weak sources tend to be either complexity-theoretic or algebraic. In this work, we focus on *algebraic sources*. That is, we consider distributions over subsets Ω which have a "nice" algebraic structure.

1.1 Algebraic Sources of Randomness

Suppose $\mathbb{F}$ is a finite field and $\Omega = \mathbb{F}^n$. The simplest class of algebraic sources is the set of *affine sources*. An affine source is simply the uniform distribution over an affine subspace $V \subseteq \mathbb{F}^n$ of dimension k. Note that since $|V| = |\mathbb{F}|^k$, the single parameter k also determines the min-entropy of the uniform distribution over the source.

Gabizon and Raz [15] constructed an explicit extractor $\mathrm{Ext} : \mathbb{F}^n \to \mathbb{F}^{k-1}$, assuming the field size is bounded from below by a large enough polynomial in n. For a large enough field size q, their construction extracts almost all of the randomness from the source and has error $\varepsilon = 1/\mathrm{poly}(q)$.

The last feature is slightly undesirable, as ideally, one would like the error to decrease exponentially with k, the dimension of the source. Such a construction was given by Bourgain, Dvir and Leeman [5], albeit their construction requires the field size to be slightly super-polynomial in n, and only works for certain fields.

Over smaller fields, constructing affine extractors for small min-entropy is a more challenging task. Further, it is possible to show that any function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is constant on some affine subspace of dimension $\Omega(\log n)$ (see, e.g., Lemma 6.7 of the arXiv version of [1]), and thus one cannot hope to extract even a single bit when the min-entropy is smaller than $\log n$ (compare this with the fact that over large fields, the Gabizon–Raz extractor works for any k).

Bourgain [4] constructed an extractor that works over $\mathbb{F}_2$ for min-entropy $k = cn$ for a small constant c. This result was slightly improved by Yehudayoff [34] and Li [22]. Li [23] then presented a much improved construction which works when the min-entropy is as small as $k = \log^C(n)$ for some constant C, which was improved by [7] to $k = \log^{1+o(1)}(n)$. However, one drawback of the last two constructions is that the error parameter ε is either constant or polynomially small, whereas one would hope for it to be exponentially small in k, as in the earlier constructions of Bourgain, Yehudayoff and Li.

There are several natural ways to generalize affine sources, but some care is needed when defining those generalizations. As we remarked earlier, for an affine subspace, the single parameter k determines its size and hence the min-entropy of the corresponding source. For more complicated algebraic sets, however, as we shall now see, there are multiple parameters controlling their "complexity," and the connection between those parameters and the min-entropy of the source is not always obvious.

Dvir, Gabizon and Wigderson [10] considered *polynomial sources*, which are defined by applying a low-degree polynomial map $P : \mathbb{F}^k \to \mathbb{F}^n$ on a uniformly random input from $\mathbb{F}^k$. (Note that affine sources are a special case of polynomial sources when the degree equals one.) They further impose the algebraic condition that the Jacobian matrix of the map is of full rank, which in particular guarantees that the min-entropy of the source is high, assuming the characteristic of the field is large enough. The field size required by the construction of [10] is $\mathrm{poly}(k, d, n)^k$.

Dvir [9] studied a different generalization called *variety sources*, which are uniform distributions over sets $V \subseteq \mathbb{F}^n$ that are the common zeros of a set of low-degree polynomials. Varieties also have an associated concept of dimension, but unlike the affine case, over finite fields having a large dimension does not guarantee by itself that the set V is large, and thus this condition must be imposed explicitly. Dvir presented two constructions. The first requires exponentially large fields and works for any dimension k. The second requires the variety to have size larger than $|\mathbb{F}|^{n/2}$, but the field size depends only polynomially on the degree d of the polynomials defining V.

Over $\mathbb{F}_2$, the situation is much more mysterious. This setting is well motivated, since it turns out that explicit constructions of extractors (or even dispersers) for varieties with various parameters would imply new circuit lower bounds. Golovnev, Kulikov and Williams [18] proved multiple such results. One is that explicit extractors for varieties of size at least $2^{\varepsilon n}$ defined by constant degree polynomials would imply lower bounds for general circuits of the form Cn for larger constants C than what is currently known. They also showed that extractors for varieties of size at least $2^{0.99n}$ defined by polynomials of degree at most $n^{0.01}$ would imply super-linear lower bounds for boolean circuits of depth $O(\log n)$, a long-standing challenge in complexity theory (see also [19]).

As for constructions over $\mathbb{F}_2$, Li and Zuckerman [21] showed how to use correlation bounds against low-degree polynomials to obtain extractors for variety sources defined by degree d polynomials for $d = O(1)$ and size at least $2^{(1-c_d)n}$ for some constant c_d that depends on d. Remscrim [30] proved that the majority function is an extractor for varieties defined by polynomials of degree at most n^α and size at least 2^{n-n^β}, assuming $\alpha + \beta < 1/2$. Thus, all the known constructions are not strong enough to imply new circuit lower bounds.

1.2 Our Results

1.2.1 Extractor for Polynomial Images of Varieties. In this paper, we study the class of *polynomial images of varieties*, which generalizes both variety sources and polynomial sources. Informally, the source is specified by a variety $V \subseteq \mathbb{F}^r$ and a polynomial map $f : V \to \mathbb{F}^n$, and a sample from the source is a random variable X computed by uniformly at random picking an element $x \in V$ and outputting $f(x)$. We would like to construct an efficient extractor $\mathrm{Ext} : \mathbb{F}^n \to \{0,1\}^m$ that has small error ε and large output length m. The largest m we can hope for is the min-entropy of the input, which is approximately $k \log q$, where $q = |\mathbb{F}|$ and k is the dimension of the variety V (see Section 4 for a definition of this notion). Our main result is a construction of an extractor with $m \approx k \log q$.

Formally defining such sources takes some care, since varieties and their associated complexity parameters are easier to define over algebraically closed fields. As in previous work, we further need to assume some natural non-degeneracy conditions on the variety V and the map f. We now describe those sources in more detail.

Polynomial images of variety sources. Let $\mathbb{F}$ be a field. For a set $h_1, \ldots, h_s \in \mathbb{F}[X_1, \ldots, X_n]$, define

$$\mathcal{L}_{h_1,\ldots,h_s,\mathbb{F}} := \{c_0 + c_1 h_1 + \cdots + c_s h_s : c_0, \ldots, c_s \in \mathbb{F}\} \subseteq \mathbb{F}[X_1, \ldots, X_n],$$

i.e., $\mathcal{L}_{h_1,\ldots,h_s,\mathbb{F}}$ is the linear span of $h_1, \ldots, h_s$ and 1 over $\mathbb{F}$.

Denote by $\overline{\mathbb{F}}$ the algebraic closure of $\mathbb{F}$. An *affine variety* $V \subseteq \overline{\mathbb{F}}^n$ over $\mathbb{F}$ is the set of common zeros of a set of polynomials in $\mathbb{F}[X_1, \ldots, X_n]$. Two parameters naturally associated with a variety V are its *dimension*, denoted $\dim V$, which equals the length of the maximal chain with respect to inclusion of distinct irreducible subvarieties, and its *degree*, denoted $\deg V$, which is the number

of intersection points of the variety with an affine subspace of codimension $\dim V$ in general position (we refer to Section 4 for more formal definitions).

Definition 1.2 ((n,k,d) algebraic source). *Let $n, d \in \mathbb{N}^+$ and $k \in \mathbb{N}$. We say a distribution D over $\mathbb{F}_q^n$ is an (n,k,d)* algebraic source over $\mathbb{F}_q$ *if there exist $r \in \mathbb{N}$, an affine variety $V \subseteq \overline{\mathbb{F}}_q^r$ over $\mathbb{F}_q$, polynomials $h_1, \ldots, h_s \in \mathbb{F}_q[X_1, \ldots, X_r]$ with $\deg h_1 \geq \cdots \geq \deg h_s$, and $f_1, \ldots, f_n \in \mathcal{L}_{h_1,\ldots,h_s,\mathbb{F}_q}$ such that $D = f(U_{V(\mathbb{F}_q)})$, where $f : \overline{\mathbb{F}}_q^r \to \overline{\mathbb{F}}_q^n$ is the polynomial map defined by $f_1, \ldots, f_n$, and $U_{V(\mathbb{F}_q)}$ is the uniform distribution over $V(\mathbb{F}_q) := V \cap \mathbb{F}_q^r$, and further, the following conditions hold:*

1. *At least one irreducible component of V of dimension $\dim V$ is absolutely irreducible.*
2. *For every irreducible component V_0 of dimension $\dim V$ that is absolutely irreducible, the dimension of $\overline{f(V_0)}$ is at least k, where $\overline{f(V_0)} \subseteq \overline{\mathbb{F}}_q^n$ denotes the closure of $f(V_0)$, i.e., the smallest affine variety over $\mathbb{F}_q$ containing $f(V_0)$.*
3. $\deg V \cdot \prod_{i=1}^{k} \deg h_i \leq d$.[1]

In addition, we say D is an irreducible (n,k,d) *algebraic source over $\mathbb{F}_q$ if V can be chosen to be irreducible. We say D is a* minimal (n,k,d) *algebraic source over $\mathbb{F}_q$ if V can be chosen to have dimension k. Finally, we say D is an* irreducibly minimal (n,k,d) *algebraic source over $\mathbb{F}_q$ if V can be chosen to be irreducible of dimension k.*

The conditions in Definition 1.2 may look a bit contrived at first glance. However, as we now explain, they are quite natural, and indeed some form of them, as observed in previous work, is necessary.

The third condition is simply a convenient way to "pack" multiple "complexity" parameters of the components of the source that arise in the analysis. That is, d is a single complexity parameter that, in particular, bounds the degree of the variety V and the product of degrees of the polynomial map f. Having d as a single parameter simplifies the statements of our theorems and clarifies the dependence between the various parameters: the larger d is, the larger the field size we require and the smaller the output length of the extractor.

The purpose of the first two conditions is to guarantee that our source has enough min-entropy. As observed in previous work [9, 10], it is quite easy to come up with simple varieties V (even of high dimension) or polynomial maps f (even of low degree) such that sources arising as $f(V)$ would have very few points in $\mathbb{F}_q^n$, so that there will be little to no randomness to extract.

The first condition is analogous to (and, as shown in the full version of this paper, roughly equivalent to) Dvir's [9] condition that the variety V contains enough points in $\mathbb{F}_p^n$. The second condition is analogous to (and, over fields of large characteristic, implied by) the full-rank Jacobian condition of Dvir, Gabizon and Wigderson [10]. Thus, not only is some form of conditions 1 and 2 necessary for proving any meaningful results, but moreover, these conditions naturally generalize the conditions imposed by previous related works.

[1] Note that $\dim \overline{f(V)} \geq k$ by previous conditions. So we necessarily have $s \geq k$ and $\deg h_i \geq 1$ for $i \in [k]$. This also implies $\deg V \leq d$.

Finally, we note that the name "(n,k,d) algebraic sources" suppresses the dependence on the parameter r in the definition, which is the ambient dimension in which the variety V lies. This is because our result, stated next, has no dependence on r. Even in the case where r is very large with respect to n, k and d, our results only depend on the latter three parameters. Further, note that when r is very large, $\dim V$ can also be very large compared with n and k. However, as the definition hints, we will reduce this case to the case where $\dim V = k$.

We can now state our main theorem.

Theorem 1. *Let $n, d \in \mathbb{N}^+$, $k \in \mathbb{N}$, and $\varepsilon \in (0, 1/2]$. Let q be a power of a prime p. Suppose $q \geq (nd/\varepsilon)^c$, where $c > 0$ is a large enough absolute constant. Then there exists an explicit ε-extractor $\mathrm{Ext} : \mathbb{F}_q^n \to \{0,1\}^m$ for (n,k,d) algebraic sources over $\mathbb{F}_q$ with output length $m \geq k \log q - 4 \log\log p - O(\log(nd/\varepsilon))$.*

It can be shown that any (n,k,d) algebraic source D over $\mathbb{F}_q$, where $q \geq (kd)^c$ for a sufficiently large constant $c > 0$, is (close to) a distribution with min-entropy at least $k \log q - O(\log d)$. Moreover, this estimate of the min-entropy is tight up to an additive term $O(\log d)$ if D is not an $(n, k+1, d)$ algebraic source over $\mathbb{F}_q$. See Lemma 7.4 and Proposition 7.5. Therefore, the extractor in Theorem 1 extracts most of the min-entropy from (n,k,d) algebraic sources. In addition, Theorem 1 works over finite fields of any characteristic, while the extractors by Dvir, Gabizon, and Wigderson [10] and Dvir [9] require large enough characteristics.

As is standard in the literature, by "explicit" we mean that the output of the extractor is computable in time $\mathrm{poly}(n, \log q)$ (note that the input length to the extractor is $n \log q$).

Along the way to proving Theorem 1, we construct several other algebraic pseudorandom objects which are interesting on their own. We mention some of these constructions when we give an overview of our construction in Section 1.3.

1.2.2 *Affine Extractors for Quasipolynomally Large Fields with Exponentially Small Error.*

Recall that an explicit affine extractor is an efficiently computable function $\mathrm{Ext} : \mathbb{F}^n \to \mathbb{F}^m$ such that for every affine subspace $V \subseteq \mathbb{F}^n$ of dimension k, and a random variable X uniformly sampled from V, $\mathrm{Ext}(X)$ is close to the uniform distribution over $\mathbb{F}^m$. We would like m to be as close to k as possible and, ideally, the error parameter ε to be exponentially small in k.

As mentioned earlier, the extractor of Gabizon and Raz [15] achieves $m = k - 1$ and error ε only polynomially small in the field size q. In particular, the error does not decrease with k. Bourgain, Dvir and Leeman [5] constructed an affine extractor with m arbitrarily close to $k/2$ and error $q^{-\Omega(k)}$. However, their construction requires q to be slightly super-polynomial in n, namely $q = n^{\Omega(\log\log n)}$, and furthermore only works for "most" prime fields $\mathbb{F}_q$. We improve the analysis of their construction and present a construction with identical parameters that works for *all* prime fields, assuming $q = n^{\Omega(\log\log n)}$.

Theorem 2. *For every $0 < \beta < 1/2$, there exists a constant C such that the following holds: Let $k \leq n$ be integers and $\mathbb{F}$ be a prime field of size $q \geq n^{C \log\log n}$. Let $m = \beta k$. There exists an efficiently computable function $E : \mathbb{F}^n \to \mathbb{F}^m$ which is an affine extractor for min-entropy k with error $q^{-\Omega(k)}$.*

1.3 Techniques

Our construction from Theorem 1 combines several techniques used in previous related constructions, as well as several new ideas which are required to successfully apply these techniques. It is convenient to think of the construction as proceeding in several steps.

Preliminary step: decomposing the sources. Our definition for algebraic sources (Definition 1.2) is quite general, and it is convenient to work with slightly "nicer" sources. We start by approximating general (n, k, d) algebraic sources as convex combinations of *irreducibly minimal* (n, k, d) algebraic sources. Recall that this means that the variety V is irreducible and has dimension k.

This step is done in Section 7: we first decompose a general source into a convex combination of irreducible sources in a manner that follows naturally from the decomposition of V itself as a union of irreducible components. We then decompose an irreducible source into irreducibly minimal sources roughly by intersecting it with a linear space of the appropriate dimension. Both parts of the arguments incur a small error.

First step: extracting a short seed. Having reduced to the case of irreducibly minimal sources, we first design an extractor that extracts a small number of bits from the source. One commonly used technique for doing that is to show that the source is an ε-biased distribution, i.e., a distribution whose nontrivial Fourier coefficients are all small. Similar methods work when the source is close to such a distribution or to a convex combination of such distributions. Analyzing and bounding the Fourier coefficients is often done using bounds on exponential sums from algebraic geometry, such as Bombieri's estimate (Theorem 4.3). We follow this general paradigm as well.

However, the case where the field characteristic is small presents some unique challenges to overcome. We first prove an extension of Bombieri's theorem for small characteristic p. This extension bounds the corresponding exponential sums save for possibly a small set of "bad" characters. Hence, we then define and study a more general class than ε-biased distributions: (ε, e)-biased distributions, which are distributions in which all but at most e of the Fourier coefficients have absolute value at most ε. We show that the sources we consider are close to convex combinations of such distributions (for meaningful values of ε and e), and construct extractors for such distributions.

Previously, the XOR lemma has been used to construct extractors for ε-biased sources; see, e.g., Rao [29]. We extend these ideas to the more general and challenging setting of (ε, e)-biased distributions. On the technical level, we construct explicit functions $f : \mathbb{F}_p^n \to \mathbb{F}_p^t$ with the following properties: for every nontrivial character ψ of $\mathbb{F}_p^t$, both the L_1 and the L_∞ norms of the Fourier transform of $\psi \circ f$ (which is a function from $\mathbb{F}_p^n$ to $\mathbb{C}$) are upper bounded by sufficiently small quantities. We in fact present two constructions of such functions f. The first is based on standard error-correcting codes over $\mathbb{F}_p$, and the second is an improved construction based on *rank-metric codes*. Those constructions appear in Section 3.2.

Second step: applying a seeded extractor. Having extracted a small number of bits, we wish to use them as a *seed* in an application of a seeded extractor on the source to extract almost all the min-entropy. The challenge, of course, is that the seed is correlated with the source, whereas a seeded extractor requires the seed to be independent of the source. Techniques for dealing with these problems were developed in [15, 16], as this is also the general methodology in their extractor constructions. This is done by analyzing the conditional distribution of the source conditioned on any possible output of the seeded extractor with a fixed seed, and showing that it maintains some nice properties. We first analyze the case where the image $f(V)$ of the polynomial map is of full rank inside $\mathbb{F}^k$, using the *effective fiber dimension theorem*. We then consider the general case. In order to reduce to that case, we apply a *rank extractor* for varieties, a notion we define and develop in this work, building upon previous work which developed rank extractors for linear spaces.

Rank extractor for varieties. Let $V \subseteq \mathbb{F}^n$ be a k-dimensional variety. We would like to obtain a map $E : \mathbb{F}^n \to \mathbb{F}^k$ which "extracts" all the rank from V, in the sense that $E(V) \subseteq \mathbb{F}^k$ is k-dimensional. The first obvious challenge is that $E(V)$ need not necessarily be a variety. It is thus natural in this case to consider the closure of $E(V)$ in $\overline{\mathbb{F}}^n$ where $\overline{\mathbb{F}}$ is the algebraic closure of $\mathbb{F}$.

Previous work has considered the case where V is a linear subspace. In this case, observe that if E is linear, then $E(V)$ is also a linear subspace. However, there clearly cannot be a single map E that preserves the dimension of all linear subspaces, as given any fixed E, one could take V to be the kernel of E. Therefore, a natural relaxation is to consider *seeded* linear rank extractors, which are collections of linear maps $E_1, \ldots, E_t$ such that for every V, most of the maps preserve the dimension. Such objects were first defined and constructed by Gabizon and Raz [15]. Improved and optimal parameters (in terms of the "seed length," i.e., the number of maps) were obtained by Forbes and Shpilka [14], and a systematic study of these objects appears in [13].

In this work, we observe that seeded linear rank extractors for extractors are also seeded linear rank extractors for varieties (see Section 5). The key insight is that rank extractors (for linear subspaces) preserve the dimensions of the tangent spaces at nonsingular points of the variety, which turns out to be a sufficient criterion.

Linear rank extractors are very useful because they enable us to condense sources that are not full-rank to full-rank sources without increasing the degrees of the polynomial maps. However, it turns out that it is also possible to construct *deterministic* rank extractors for varieties, which we do in Section 6. Such extractors are obviously not linear maps, although in our constructions, they are polynomials of fairly small degrees (polynomial in n and in the degree d of the variety). We remark that Dvir [9] constructed such an extractor for one-dimensional varieties, and his extractor is a polynomial of degree exponential in n. In addition, Dvir, Gabizon and Wigderson [10] constructed rank extractors for polynomial sources using a different technique.

Our construction adapts the construction of Dvir, Kollár and Lovett [11], who constructed different pseudorandom objects called *variety evasive sets*. By modifying their proof, we are able to show that a similar construction yields a deterministic rank extractor for varieties. This essentially follows because their map φ satisfies the property that for every low-degree variety V and every point

$b \in \mathbb{F}^k$, the intersection $\varphi^{-1}(b) \cap V$ is a finite set. Dvir, Kollár and Lovett prove it only for the case $b = \mathbf{0}$, but it is not hard to extend it to general b.

Explicit Noether normalization lemmas. As a by-product of the above construction of deterministic rank extractors for varieties, we prove explicit *Noether normalization lemmas* for affine varieties and affine algebras. The Noether normalization lemma [26, 28] is a classical result in commutative algebra and algebraic geometry, which states that any affine variety of dimension k admits a surjective *finite morphism* to an affine space of dimension k. We show that the construction in [11] in fact gives a direct construction of such a finite morphism. In contrast, the textbook proof of Nagata [26] is iterative and uses polynomials of degrees that are at least doubly exponential in the number of steps of the iteration.

Our proof is inspired by a geometric argument of Kollár, Rónyai and Szabó [20]. See Section 11 and the full version of this paper for more details.

Affine extractors with exponentially small error. Our proof of Theorem 2 follows a very similar route to the proof of the main theorem of Bourgain, Dvir and Leeman [5], who constructed such an extractor for prime fields $\mathbb{F}_q$ for "typical" primes q. Our main contribution is an improved number-theoretic lemma (Proposition 10.1) which shows how to find n distinct integers $d_1, \ldots, d_n$ with desirable number theoretic properties. The proof then proceeds by estimating the Fourier coefficient of the distribution obtained by applying our extractor to a linear subspace using an exponential sum estimate of Deligne, much in the same way as [5].

1.4 Comparison with Previous Work

The two works closest to ours are by Dvir [9] and Dvir, Gabizon and Wigderson [10], both of which construct extractors for sources with algebraic structures.

As mentioned earlier, Dvir, Gabizon and Wigderson [10] study *polynomial sources*, defined by picking an element $x \in \mathbb{F}_q^k$ uniformly at random and applying a polynomial map $f : \mathbb{F}_q^k \to \mathbb{F}_q^n$ of degree at most d. This is a special case of the sources we consider when the variety V is taken to be $\mathbb{F}_q^k$.

They further add the non-degeneracy condition that the *Jacobian* of the mapping f, namely, its matrix of partial derivatives, has full rank. This in particular guarantees that the source has a high enough min-entropy. Their main theorem gives an explicit extractor that outputs a constant fraction of the min-entropy over prime fields $\mathbb{F}_p$ of cardinality $\mathrm{poly}(n, d)^{Ck}$ for some constant C. Our construction in Theorem 1, on the other hand, works for a larger class of sources, outputs almost all the min-entropy, and works over finite fields of small characteristics as well.

Dvir [9] considers *variety sources*, which he defines as uniform distributions over sets of the type

$$\{x : f_1(x) = f_2(x) = \cdots = f_t(x) = 0\}$$

in $\mathbb{F}_p^n$, where $\deg f_i \le d$ for all i. These sources are also a special case of the type of sources we consider. One should note, however, the different usage of the term "degree" in our definitions: Dvir always refers to the degree $\deg f_i$ of the polynomials which define the variety V, whereas we refer to the degree $\deg V$ of V as an affine variety, which is often much larger.

Assuming $\dim V = k$ and $|V| \ge p^{k-c}$ for some small constant $c > 0$, Dvir's extractor [9] outputs a constant fraction of the min-entropy over prime fields of characteristic $p > d^{Cn^2}$ for some constant C. Again, Dvir uses the parameter d differently than we do in Theorem 1. In particular, in our construction, the field size q is only polynomial in the parameter d (but d might be exponential in n).

As mentioned in the discussion after Definition 1.2, our assumptions are weaker than those of [10] and [9]. Thus, as our sources is more general, the characteristic in our results can be arbitrary, and our conclusions are stronger (since we extract more output bits), it follows that in particular our result subsumes the extractors of [10] and [9].

Dvir [9] also presents a different construction that outputs a very small number of bits from very large varieties over small fields. This construction is incomparable with our results.

On the more technical level, we discuss a particular feature of our proof that distinguishes it from [9, 10] and, in particular, allows us to extend the output length.

For simplicity, consider the case of $(1, 1, d)$ algebraic sources. As mentioned in Section 1.3, we first prove an extension of Bombieri's estimate that holds even if the characteristic p is small: if p is small, this result implies that a $(1, 1, d)$ algebraic source D over $\mathbb{F}_q$ is a convex combination of (ε, d)-biased sources. That is, we allow a few large Fourier coefficients. Then we use the machinery developed in Section 3 to extract randomness from D. On the other hand, if p is large enough, then D has no large nontrivial Fourier coefficients; it is ε-biased. In this case, the XOR lemma is sufficient, as argued in [9, 10].

To apply Bombieri's estimate to a high-dimensional affine variety V, we follow [9, 10] and decompose V into a family of affine curves C_i such that the polynomial f that does not vanish identically on V still does not vanish on most C_i.

In [10], this is achieved using an argument based on the Jacobian criterion for algebraic independence, but it works only when the characteristic p is large. Instead of using this argument, we use the decomposition of (n, k, d) algebraic sources into irreducibly minimal (n, k, d) algebraic sources proved in Section 7, whose proof is based on the effective fiber dimension theorem (Theorem 4.6) and works for any characteristic.

The last idea we introduce is the use of the effective Lang–Weil bound (Theorem 4.2), which allows us to extract almost $\log q$ bits. To explain the idea, consider an affine variety $V \subseteq \mathbb{A}^n_{\mathbb{F}_q}$ and write $V(\mathbb{F}_q)$ as a disjoint union of $C_i(\mathbb{F}_q)$ for a family of affine curves C_i over $\mathbb{F}_q$. Let f be a low-degree polynomial and assume for simplicity that f is non-constant on every C_i. Let χ be a nontrivial character of $\mathbb{F}_q$. The following win-win argument was used in [10] to bound the bias $\delta := \left|\mathbb{E}_{x \in V(\mathbb{F}_q)}[\chi(f(x))]\right|$: For a curve C_i, if $|C_i(\mathbb{F}_q)|$ is small, say $|C_i(\mathbb{F}_q)| \le \Delta$ for some threshold Δ, then its contribution to the bias δ is small assuming that V has many rational points. On the other hand, if $C_i(\mathbb{F}_q) > \Delta$, then Bombieri's estimate (Lemma 8.3),

together with the fact that

$$\left|\mathop{\mathbb{E}}_{x\in C_i(\mathbb{F}_q)}[\chi(f(x))]\right| = \frac{\left|\sum_{x\in C_i(\mathbb{F}_q)}[\chi(f(x))]\right|}{|C_i(\mathbb{F}_q)|} \le \frac{\left|\sum_{x\in C_i(\mathbb{F}_q)}[\chi(f(x))]\right|}{\Delta},$$

implies that $\left|\mathbb{E}_{x\in C_i(\mathbb{F}_q)}[\chi(f(x))]\right|$ is small. Considering all curves C_i shows that the bias is small. We note that no information about $|C_i(\mathbb{F}_q)|$ was used in this win-win argument. For this reason, the choice of threshold Δ cannot be too large or too small, and the resulting extractors only extract a constant fraction of $\log q$ bits. To improve the output length, we observe that the effective Lang–Weil bound (Theorem 4.2) together with gives more information about $|C_i(\mathbb{F}_q)|$. In particular, for an irreducible affine curve C, the number $|C(\mathbb{F}_q)|$ is either close to q or very small, depending on whether C is absolutely irreducible. Exploiting this fact yields an explicit construction of deterministic extractors that output almost $\log q$ bits.

1.5 Open Problems

While improving the dependence on any of the parameters in our construction remains an open problem, in our opinion, the main challenge is reducing the field size. In our construction for polynomial images of varieties (Theorem 1), we require field size $\mathrm{poly}(n, 1/\varepsilon, d)$. We stress that for certain varieties, d can be exponential in n (although it is by no means necessarily so). Can we construct extractors for significantly smaller fields, perhaps even constant size?

As mentioned above, over very small fields, such as $\mathbb{F}_2$, certain Ramsey-theoretic lower bounds imply that constructions such as ours that work for any min-entropy cannot exist. A key reason to study $\mathbb{F}_2$ is that explicit extractors with certain parameters imply new circuit lower bounds.

In our construction of new affine extractors (Theorem 2), we obtain a field size that is slightly super-polynomial in n. It is a very appealing open problem to reduce the field size to a polynomial in n.

A related problem is reducing the degree of our deterministic rank extractor. In Section 6, we construct a deterministic rank extractor for varieties whose degree is $\mathrm{poly}(n, d)$ for degree d varieties. Reducing the degree, perhaps to depend only on d, would help lower the field size requirement for the extractor for polynomial images of varieties to depend only on the degree.

We end with two general questions. Can our constructions or techniques help in designing extractors for larger and more general classes of sources, either algebraic or complexity-theoretic? Do our constructions have any complexity-theoretic implications, such as lower bounds for certain models of computation?

2 NOTATIONS AND PRELIMINARIES

Let $\mathbb{N} = \{0, 1, \dots\}$, $\mathbb{N}^+ = \{1, 2, \dots\}$, and $[n] = \{1, 2, \dots, n\}$ for $n \in \mathbb{N}$. Write $\mathbb{Z}_n$ for the cyclic group $\{0, 1, \dots, n-1\}$ with addition modulo n.

The cardinality of a set S is denoted by $|S|$. We also use $|c|$ to denote the absolute value of a number $c \in \mathbb{C}$. Denote by $\log x$ the base 2 logarithm of x, and by $\ln x$ the natural logarithm of x. For sets A and B, denote by $A \setminus B$ the set difference $\{x \in A : x \notin B\}$. The restriction of a map $f : A \to B$ to a subset $A' \subseteq A$ is denoted by $f|_{A'}$, which is a map from A' to B.

We write $x \sim D$ if x is sampled from a distribution D. The *support* of a distribution D over a finite set Ω is $\mathrm{supp}(D) := \{a \in \Omega : \Pr[D = a] \neq 0\}$. For an event A that occurs with a nonzero probability under a distribution D, write $D|_A$ for the distribution of D conditioned on A. The product distribution of two distributions D, D' is denoted by $D \times D'$. The *statistical distance* between two distributions D, D' over a finite set Ω is defined to be

$$\Delta(D, D') := \max_{A\subseteq\Omega} |\Pr[D \in A] - \Pr[D' \in A]|.$$

Two distributions D and D' are *ε-close* if their statistical distance is at most ε, and we write $D =_\varepsilon D'$ for this statement.

The uniform distribution over a finite set S is denoted by U_S. For $n \in \mathbb{N}$, denote by U_n the uniform distribution over $\{0, 1\}^n$.

The *min-entropy* of a distribution D over a finite set Ω is

$$H_{\min}(D) := -\log(\max_{a\in\Omega} \Pr[D = a]).$$

We say D is a *k-source* if $H_{\min}(D) \ge k$.

Let Ω and B be finite sets, and let $\mathcal{D}$ be a class of distributions over Ω. A function $\mathrm{Ext} : \Omega \to B$ is said to be a *(deterministic) ε-extractor* for $\mathcal{D}$ if $\mathrm{Ext}(D) =_\varepsilon U_B$ for all $D \in \mathcal{D}$. A function $\mathrm{Ext} : \Omega \times \{0,1\}^\ell \to B$ is said to be a *seeded ε-extractor* for $\mathcal{D}$ if $\mathrm{Ext}(D \times U_\ell) =_\varepsilon U_B$ for all $D \in \mathcal{D}$, where $\ell \in \mathbb{N}$ is called the *seed length* of Ext.

3 SOURCES WITH LOW BIAS AND THEIR EXTRACTORS

We consider several natural extensions of ε-biased sources which are useful for our extractor constructions. We then show how to extract randomness from such sources.

3.1 (ε, e)-Biased Sources

Let A be a finite abelian group and let $\widehat{A}$ denote the dual group of A, that is, the group of characters over A. A distribution D over A is *ε-biased* if $|\mathbb{E}[\chi(D)]| \le \varepsilon$ for all nontrivial characters $\chi \in \widehat{A}$. This is a standard definition, introduced in [27], which has been immensely useful in the construction of extractors and in the theory of pseudorandomness in general.

We now introduce two natural generalizations. We say D is *(ε, e)-biased* if $|\mathbb{E}[\chi(D)]| \le \varepsilon$ for all but at most e characters $\chi \in \widehat{A}$. And we say D is *strongly (ε, e)-biased* if the set of $\chi \in \widehat{A}$ satisfying $|\mathbb{E}[\chi(D)]| > \varepsilon$ is contained in an abelian subgroup of A of size at most e. The usefulness of the latter definition will be clear shortly.

Suppose that A and B are finite groups. We wish to bound the bias of conditional distributions over A (or B), assuming bounds on the bias of a distribution over $A \times B$. We bound the bias of the marginal distribution D_2 conditioned on any value of D_1.

Corollary 3.1. *Let A and B be finite abelian groups. Identify $\widehat{A} \times \widehat{B}$ with $\widehat{A \times B}$ so that $(\chi, \theta)(x, y) = \chi(x)\theta(y)$ for $(x, y) \in A \times B$ and $(\chi, \theta) \in \widehat{A} \times \widehat{B}$. Let $D = (D_1, D_2)$ be a joint distribution over $A \times B$. Let $\varepsilon, \varepsilon' > 0$. Assume that every character $\chi \in \widehat{A \times B} \cong \widehat{A} \times \widehat{B}$*

satisfying $\mathbb{E}[\chi(D)] > \varepsilon$ is contained in the subgroup $\widehat{A} \times \{1\}$. Then with probability at least $1-\varepsilon'$ over $x \sim D_1$, the conditional distribution $D_2|_{D_1=x}$ is $|A|\varepsilon/\varepsilon'$-biased.

3.2 Extraction via the XOR Lemma and Rank-Metric Codes

In the full version of the paper, we construct extractors for ε-biased sources, (ε, e)-biased sources and strongly (ε, e)-biased sources. For ε-biased sources we use known constructions that extract randomness using the XOR lemma as in [29]. Over large characteristic we use constructions based on rank metric codes.

The following construction allows us to extract randomness from ε-biased sources over $\mathbb{F}_q$. We use it for fields of large characteristic.

Lemma 3.2 ([29, Lemma 4.4]). *Let $f : \mathbb{Z}_N \to \mathbb{Z}_M$ be the map sending $a \bmod N$ to $a \bmod M$ for $a \in \{0, 1, \ldots, N-1\}$. Let ψ be a character of $\mathbb{Z}_M$. Then $\left\|\widehat{\psi \circ f}\right\|_1 \le c \log N$, where c is an absolute constant.*

When p is large but $\mathbb{F}_q$ is possibly non-prime, we simply apply the mod-M function to the last $\mathbb{F}_p$-coordinate of $\mathbb{F}_q$ and use the following corollary of Lemma 3.2.

Corollary 3.3. *Let $f : \mathbb{Z}_N^t \to \mathbb{Z}_N^{t-1} \times \mathbb{Z}_M$ be the map that sends $(a_1, \ldots, a_{t-1}, a \bmod N)$ to $(a_1, \ldots, a_{t-1}, a \bmod M)$ for every $(a_1, \ldots, a_{t-1}, a) \in \mathbb{Z}_N^{t-1} \times \{0, 1, \ldots, N-1\}$. Let ψ be a character of $\mathbb{Z}_N^{t-1} \times \mathbb{Z}_M$. Then $\left\|\widehat{\psi \circ f}\right\|_1 \le c \log N$, where c is an absolute constant.*

Lemma 3.4. *Let $f : \mathbb{Z}_N^t \to \mathbb{Z}_N^{t-1} \times \mathbb{Z}_M$ be the map in Corollary 3.3. Then for every ε-biased distribution D over $\mathbb{Z}_N^t$, $f(D)$ is ε'-close to the uniform distribution over $\mathbb{Z}_N^{t-1} \times \mathbb{Z}_M$, where $\varepsilon' = \varepsilon \cdot (N^{t-1}M)^{1/2} \cdot c \log N + M/N$ and c is an absolute constant.*

The XOR lemma requires the distribution to be ε-biased. However, when the characteristic is small, we need to deal with the more general class of (ε, e)-biased distributions, where e is small. In the full version of the paper we prove the following theorem.

Theorem 3.5. *Let n, t, e be positive integers and $\varepsilon, \varepsilon' \in (0, 1)$. Let $n' = \min\{\lfloor 2\log_p(1/\varepsilon) - 2\log_p(16e/\varepsilon'^2)\rfloor, n\}$. Suppose $t \le n' - 3 - 2\log_p(2e/\varepsilon')$. Then there exists an explicit ε'-extractor $\mathsf{Ext} : \mathbb{F}_p^n \to \mathbb{F}_p^t$ for strongly (ε, e)-biased sources.*

4 PRELIMINARIES ON ALGEBRAIC GEOMETRY

We refer to section 4 of the full version of our paper for preliminaries and notations on algebraic geometry that we require. One can also refer to a standard text, e.g., [31, 33]. In this condensed version, we simply cite a few of the claims we need for later sections.

Theorem 4.1 (Fiber dimension theorem). *Suppose $\varphi : V \to V'$ is a dominant morphism between irreducible affine varieties over an algebraically closed field $\mathbb{F}$. Then for every $b \in \varphi(V)$ and every irreducible component Z of $\varphi^{-1}(b)$, it holds that*

$$\dim Z \ge \dim V - \dim V'.$$

Moreover, there exists $U \subseteq \varphi(V)$ such that U is a dense open subset of V' and $\dim \varphi^{-1}(b) = \dim V - \dim V'$ holds for all $b \in U$.

See, e.g., [31, §I.6.3, Theorem 7] for a proof.

Theorem 4.2 (Effective Lang–Weil bound). *Let $V \subseteq \mathbb{A}^n_{\mathbb{F}_q}$ be an absolutely irreducible affine variety over $\mathbb{F}_q$ of dimension k and degree d. Then*

$$|V(\mathbb{F}_q) - q^k| < (d-1)(d-2)q^{k-1/2} + 5d^{13/3}q^{k-1}.$$

In particular, we have $|V(\mathbb{F}_q)| \ge q^k/2$ if $q \ge 20d^5$.

Bombieri's estimate for exponential sums. Bombieri's estimate gives an upper bound for exponential sums over rational points of curves over $\mathbb{F}_q$.

Theorem 4.3 ([3, Theorem 6]). *Let $C \subseteq \mathbb{A}^n_{\mathbb{F}_q}$ be an affine curve of degree d_1 over a finite field $\mathbb{F}_q$ of characteristic p. Let $\sigma : \mathbb{F}_p \to \mathbb{C}^\times$ be the character $x \mapsto e^{2\pi i x/p}$ of $\mathbb{F}_p$. Suppose $f \in \mathbb{F}_q[X_1, \ldots, X_n]$ is a polynomial of degree d_2 such that for any $g \in \overline{\mathbb{F}}_q[X_1, \ldots, X_n]$ and any irreducible component C_0 of C, the function $f - (g^p - g)$ does not vanish identically on C_0. Then*

$$\left|\sum_{x \in C(\mathbb{F}_q)} (\sigma \circ \mathrm{Tr} \circ f)(x)\right| \le (d_1^2 + 2d_1d_2 - 3d_1)q^{1/2} + d_1^2.$$

where Tr denotes the trace map from $\mathbb{F}_q$ to $\mathbb{F}_p$.

Noether normalization. The Noether normalization lemma, due to Noether [28] states that an affine variety V of dimension k over an infinite field $\mathbb{F}$ admits a finite morphism $\varphi : V \to \mathbb{A}^k_{\mathbb{F}}$. Moreover, φ may be chosen to be a linear map. We give the following quantitative version of this result, which states that the coefficients that specify the linear map can be chosen from a finite subset $S \subseteq \mathbb{F}$ provided that S is large enough.

Lemma 4.4 (Noether normalization). *Let $V \subseteq \mathbb{A}^n_{\mathbb{F}}$ be an affine variety of dimension k and degree d over a field $\mathbb{F}$. Suppose S is a finite subset of $\mathbb{F}$ of size greater than d. Then there exists a polynomial map $\varphi : \mathbb{A}^n_{\mathbb{F}} \to \mathbb{A}^k_{\mathbb{F}}$ defined by linear polynomials $\ell_i = \sum_{j=1}^n c_{i,j}X_i \in \mathbb{F}[X_1, \ldots, X_n]$ with coefficients $c_{i,1}, \ldots, c_{i,n} \in S$ for $i = 1, \ldots, k$ such that $\varphi|_V : V \to \mathbb{A}^k_{\mathbb{F}}$ is a finite morphism.*

For convenience, we also prove the following lemma, which guarantees the existence of linear polynomials achieving simultaneous Noether normalization for two affine varieties.

Lemma 4.5. *Let $\mathbb{K}_1$ and $\mathbb{K}_2$ be extension fields of a field $\mathbb{F}$. For $i = 1, 2$, let $V_i \subseteq \mathbb{A}^n_{\mathbb{K}_i}$ be an affine variety of dimension k_i and degree d_i over $\mathbb{K}_i$. Suppose S is a finite subset of $\mathbb{F}$ of size greater than $d_1 + d_2$. Then there exist linear polynomials $\ell_1, \ldots, \ell_{\max\{k_1,k_2\}} \in \mathbb{F}[X_1, \ldots, X_n]$ with coefficients in S such that the morphism $V_i \to \mathbb{A}^{k_i}_{\mathbb{K}_i}$ defined by $\ell_1, \ldots, \ell_{k_i}$ is finite for $i = 1, 2$.*

Effective fiber dimension theorem. We also need an effective version of the fiber dimension theorem. To suit our needs, we first formulate the theorem in the following general form. Recall that for $h_1, \ldots, h_s \in \mathbb{F}[X_1, \ldots, X_n]$, we denote by $\mathcal{L}_{h_1,\ldots,h_s,\mathbb{F}}$ the linear span of $h_1, \ldots, h_s$ and 1 over $\mathbb{F}$.

Theorem 4.6 (Effective fiber dimension theorem – general form). *Let $V \subseteq \mathbb{A}^n$ be an irreducible affine variety of dimension k over an algebraically closed field $\mathbb{F}$. Let $h_1, \ldots, h_s \in \mathbb{F}[X_1, \ldots, X_n]$*

with $\deg h_1 \geq \cdots \geq \deg h_s$. *Let* $f_1, \ldots, f_m \in \mathcal{L}_{h_1,\ldots,h_s,\mathbb{F}}$, *which define a polynomial map* $f : \mathbb{A}^n \to \mathbb{A}^m$. *Let* $k' = \dim \overline{f(V)}$.

Let $j_1, \ldots, j_{k'} \in [m]$ *such that the morphism* $f' : V \to \mathbb{A}^{k'}$ *defined by* $f_{j_1}, \ldots, f_{j_{k'}}$ *is dominant. Let* $V_{f'} \subseteq \mathbb{A}^n_{\mathbb{F}(Y_1,\ldots,Y_{k'})}$ *be the generic fiber of* f'. *Finally, let* $\ell_1, \ldots, \ell_k \in \mathbb{F}[X_1, \ldots, X_n]$ *be linear polynomials such that both the morphism* $\pi : V \to \mathbb{A}^k$ *defined by* $\ell_1, \ldots, \ell_k$ *and the morphism* $\tau : V_{f'} \to \mathbb{A}^{k-k'}_{\mathbb{F}(Y_1,\ldots,Y_{k'})}$ *defined by* $\ell_1, \ldots, \ell_{k-k'}$ *are finite.*

Let $t \in \{0, \ldots, k - k'\}$. *Then there exists a polynomial* $P \in \mathbb{F}[X_1, \ldots, X_n]$ *of degree at most* $k' \cdot \deg V \cdot \prod_{i=1}^{k'} \deg h_i$ *that does not vanish identically on* V *such that the following holds: Let* $\varphi : \mathbb{A}^n \to \mathbb{A}^{t+m}$ *be the polynomial map defined by* $\ell_1, \ldots, \ell_t, f_1, \ldots, f_m$. *Then for every* $a \in V$ *satisfying* $P(a) \neq 0$, *the fiber* $\varphi|_V^{-1}(\varphi(a))$ *is equidimensional of dimension* $k - k' - t$.

As a corollary, we have the following effective fiber dimension theorem, stated in a more standard form.

Corollary 4.7 (Effective fiber dimension theorem – standard form). *Let* $V \subseteq \mathbb{A}^n$ *be an irreducible affine variety over an algebraically closed field* $\mathbb{F}$. *Let* $h_1, \ldots, h_s \in \mathbb{F}[X_1, \ldots, X_n]$ *with* $\deg h_1 \geq \cdots \geq \deg h_s$. *Let* $f_1, \ldots, f_m \in \mathcal{L}_{h_1,\ldots,h_s,\mathbb{F}}$, *which define a polynomial map* $f : \mathbb{A}^n \to \mathbb{A}^m$. *Finally, let* $W = \overline{f(V)} \subseteq \mathbb{A}^m$. *Then there exists a polynomial* $P \in \mathbb{F}[X_1, \ldots, X_n]$ *of degree at most* $\dim W \cdot \deg V \cdot \prod_{i=1}^{\dim W} \deg h_i$ *that does not vanish identically on* V *such that for every* $a \in V$ *satisfying* $P(a) \neq 0$, *the fiber* $f|_V^{-1}(f(a))$ *is equidimensional of dimension* $\dim V - \dim W$.

Degree bound for the images of affine varieties. Finally, we need the following degree bound for the images of affine varieties (or more precisely, their closures) under polynomial maps.

Lemma 4.8. *Let* $V \subseteq \mathbb{A}^n_{\mathbb{F}}$ *be an affine variety over a field* $\mathbb{F}$. *Let* $h_1, \ldots, h_s \in \mathbb{F}[X_1, \ldots, X_n]$ *with* $\deg h_1 \geq \cdots \geq \deg h_s$. *Let* $f_1, \ldots, f_m \in \mathcal{L}_{h_1,\ldots,h_s,\mathbb{F}}$, *which define a polynomial map* $f : \mathbb{A}^n_{\mathbb{F}} \to \mathbb{A}^m_{\mathbb{F}}$. *Finally, let* $W = \overline{f(V)} \subseteq \mathbb{A}^m_{\mathbb{F}}$. *Then*

$$\deg W \leq \deg V \cdot \prod_{i=1}^{\dim W} \deg h_i.$$

5 LINEAR SEEDED RANK EXTRACTORS FOR VARIETIES

In this section, we consider the problem of constructing *seeded rank extractors for varieties* that are linear: i.e., a set of *linear* maps such that for every variety V most of the maps in the set preserve the dimension of V. We show that these objects are simply linear *seeded rank extractors for subspaces*, a well-known linear algebraic pseudorandom object for which explicit constructions were given in [12, 14, 15].

The proof is based on the notion of *tangent spaces* of varieties, which are linear subspaces that are local first-order approximations of varieties. Intuitively, for an affine variety V, as we look at smaller and smaller neighborhoods of a *nonsingular point* a of V, the tangent space T_aV would become a better and better approximation of V. Thus, one should expect that a linear map that preserves the dimension of T_aV, which is a subspace, also preserves the dimension of V. While it is not entirely obvious what "smaller and smaller neighborhoods" mean in the Zariski topology, we will see that the claim is indeed true and follows from general facts in algebraic geometry.

Fix $\mathbb{F}$ to be an algebraically closed field throughout this section. We first formally define seeded rank extractors for varieties and subspaces.

Definition 5.1 (Seeded rank extractors). *Let* $\varphi_1, \ldots, \varphi_\ell : \mathbb{A}^n \to \mathbb{A}^m$ *be polynomial maps, where* $n \geq m$. *We say* $(\varphi_i)_{i\in[\ell]}$ *is an* (n, m, k, ε) seeded rank extractor for varieties *(resp. subspaces) if for every affine variety (resp. linear subspace)* $V \subseteq \mathbb{A}^n$ *over* $\mathbb{F}$ *of dimension at least* k, *all but at most* ε*-fraction of* φ_i *satisfy* $\dim \overline{\varphi_i(V)} = m$ *(or equivalently,* $\varphi_i|_V : V \to \mathbb{A}^m$ *is dominant). We call* $\log \ell$ *the* seed length *of the seeded rank extractor.*

In addition, we say $(\varphi_i)_{i\in[\ell]}$ *is* linear *if each* φ_i *is a linear map, i.e., defined by linear polynomials.*

The optimal choice of k is $k = m$, in which case the seeded rank extractor is "lossless." Explicit linear (n, m, k, ε) seeded rank extractors for subspaces with seed length $O(\log n + \log(1/\varepsilon))$ and $k = m$ was first constructed by Gabizon and Raz [15]. We use an improved construction given in [12, 14].

Lemma 5.2 ([12, 14]). *Let* $n \in \mathbb{N}^+$ *and* $m \in [n]$. *Let* $\omega \in \mathbb{F}^\times$ *such that the multiplicative order of* ω *is at least* n. *Let* $s_1, \ldots, s_\ell$ *be distinct elements in* $\mathbb{F}^\times$. *For* $i \in [\ell]$, *let* $\varphi_i : \mathbb{A}^n \to \mathbb{A}^m$ *be the linear map defined by the* $m \times n$ *matrix* $((\omega^{j'-1}s_i)^{j-1})_{j'\in[m],j\in[n]}$. *In other words,* φ_i *maps* $(a_1, \ldots, a_n)$ *to*

$$\left(\sum_{j=1}^{n} s_i^{j-1} a_j, \sum_{j=1}^{n} (\omega s_i)^{j-1} a_j, \ldots, \sum_{j=1}^{n} (\omega^{m-1} s_i)^{j-1} a_j\right).$$

Then $(\varphi_i)_{i\in[\ell]}$ *is a linear* (n, m, m, ε) *seeded rank extractor for subspaces, where* $\varepsilon = m(n - m)/\ell$.

The main result of this section is the following theorem.

Theorem 5.3. *An* (n, m, k, ε) *linear seeded rank extractor for subspaces is also an* (n, m, k, ε) *linear seeded rank extractor for varieties.*

Corollary 5.4. *The construction* $(\varphi_i)_{i\in[\ell]}$ *in Lemma 5.2 is a linear* (n, m, m, ε) *seeded rank extractor for varieties, where* $\varepsilon = m(n - m)/\ell$.

The proof of Theorem 5.3 appears in the full version of the paper.

6 DETERMINISTIC RANK EXTRACTORS FOR VARIETIES

Let $\mathbb{F}$ be an algebraically closed field. In this section, we consider the problem of constructing explicit *deterministic (lossless) rank extractors/condensers for varieties*. These are polynomial maps $\mathbb{A}^n \to \mathbb{A}^m$ that preserve the dimension of low-degree affine varieties $V \subseteq \mathbb{A}^n$ over $\mathbb{F}$ but reduce the dimension of the ambient space.

Dvir, Gabizon and Wigderson [10] constructed explicit deterministic rank extractors for *polynomial sources*. These objects can also be viewed as deterministic rank extractors for varieties that are the closures of the images of polynomial maps. A key technique used in their analysis is the *Jacobian criterion for algebraic independence*, which requires the characteristic of $\mathbb{F}$ to be zero or large.

To solve the problem for general varieties, one natural approach is generalizing the Jacobian criterion for algebraic independence. A

key step in the proof of [10] is showing that a certain polynomial associated with the Jacobian matrix is nonzero. Thus, it is natural for us to show that a similar polynomial does not vanish completely on affine varieties and that this is sufficient for constructing deterministic rank extractors for varieties.

While this idea can be made rigorous, the problem is that proving the nonvanishing of a polynomial on an affine variety appears to be challenging. We need to show that not only is the polynomial nonzero, but it remains nonzero modulo the ideal defining the variety. It is not clear to us how to prove such a result due to the generality of the variety.

The DKL construction. Instead of using a Jacobian-based construction, we take a different approach. Namely, we show that the explicit construction of *variety evasive sets* by Dvir, Kollár, and Lovett [11] can be used to construct deterministic rank extractors for varieties. Variety evasive sets are large finite subsets of $\mathbb{A}^n$ that have small intersections with varieties of low degree and low dimension. While they do not give deterministic rank extractors for varieties in general, we show that the construction of variety evasive sets in [11] does give such a construction.

More specifically, Dvir, Kollár and Lovett [11] construct explicit variety evasive sets by constructing an explicit polynomial map $\varphi : \mathbb{A}^n \to \mathbb{A}^m$ defined by polynomials $f_1, \ldots, f_m \in \mathbb{F}[X_1, \ldots, X_n]$ such that the intersection of $\varphi^{-1}(\mathbf{0}) = V(f_1, \ldots, f_m)$ with any low-degree variety of dimension at most m is finite, where $\mathbf{0}$ denotes the origin of $\mathbb{A}^n$. We observe that this remains true if $\varphi^{-1}(\mathbf{0})$ is replaced by $\varphi^{-1}(b)$ for any $b \in \mathbb{A}^m$. In other words, for any low-degree variety V of dimension at most m, the polynomial map φ restricts to a morphism $\varphi|_V : V \to \mathbb{A}^m$ whose fibers are all finite sets. In the terminology of algebraic geometry, this means $\varphi|_V$ is a *quasi-finite morphism*. By the fiber dimension theorem (Theorem 4.1), we then have $\dim \overline{\varphi(V)} = \dim(V)$.

In this section, we construct explicit deterministic rank extractors and rank condensers for varieties by adapting the analysis in [11]. We also formulate the construction in a way that highlights the connection with linear error-correcting codes. In particular, a linear MDS code yields a deterministic rank extractor for varieties in the sense that the coefficients of the polynomials that define the rank extractor are specified by a parity-check matrix of the code.

In Section 11, we will show that the polynomial map φ has the stronger property that $\varphi|_V$ is a *finite morphism*, not just quasi-finite, and this gives explicit Noether normalization lemmas for affine varieties and affine algebras.

Our Explicit Construction. We first define deterministic rank extractors and rank condensers for varieties.

Definition 6.1 (Deterministic rank extractors/condensers for varieties). *Let $n \in \mathbb{N}^+$ and $m \in [n]$. A polynomial map $\varphi : \mathbb{A}^n \to \mathbb{A}^m$ is an (n, m, k, d)* deterministic (lossless) rank condenser *if $\dim \overline{\varphi(V)} = \dim V$ for every affine variety $V \subseteq \mathbb{A}^n$ over $\mathbb{F}$ of dimension at most k and degree at most d. When $k = m$, we also say φ is an (n, m, d)* deterministic (lossless) rank extractor.

k-regular matrices. Let $n \in \mathbb{N}^+$ and $m, k \in [n]$. We say a matrix $M \in \mathbb{F}^{m \times n}$ is *k-regular* if any k distinct columns of M are linearly independent. (The same definition was given in [11] but for only for the special case where $k = m$.)

The following lemma gives a coding-theoretic characterization of k-regularity. Its proof is straightforward.

Lemma 6.2. *Let $\mathbb{K}$ be a subfield of $\mathbb{F}$ and let $M \in \mathbb{K}^{m \times n} \subseteq \mathbb{F}^{m \times n}$, where $n \in \mathbb{N}^+$ and $m, k \in [n]$. The following statements hold.*

- *M is k-regular iff there does not exist a nonzero vector $u \in \mathbb{K}^n$ of Hamming weight at most k such that $Mu = 0$.*
- *Suppose $k = m$. Then M is k-regular iff it is an* MDS matrix*, i.e., every maximal minor of M is nonzero.*

In particular, assuming $\mathbb{K}$ is a finite field, the matrix M is k-regular iff the linear code $C = \{u \in \mathbb{K}^n : Mu = 0\}$ over $\mathbb{K}$ defined by the parity check matrix M has minimum distance at least $k + 1$. And if $k = m$, then M is k-regular iff C is a linear MDS code of minimum distance $k + 1$, i.e., it is a linear code of dimension $n - k$ and minimum distance $k + 1$.[2]

The construction. We now present the explicit construction of deterministic rank extractors and condensers for varieties. It is based on the explicit construction of variety evasive sets in [11].

Let $n, d \in \mathbb{N}^+$ and $m, k \in [n]$. Let $d_1, \ldots, d_n$ be n pairwise coprime integers greater than d.[3] Let $M = (c_{i,j})_{i \in [m], j \in [n]} \in \mathbb{F}^{m \times n}$ be a k-regular matrix. Let $\varphi = \varphi(M) : \mathbb{A}^n \to \mathbb{A}^m$ be the polynomial map

$$\varphi : (a_1, \ldots, a_n) \mapsto \left(\sum_{j=1}^{n} c_{1,j} a_j^{d_j}, \ldots, \sum_{j=1}^{n} c_{m,j} a_j^{d_j} \right).$$

We remark that, curiously, the construction above is very similar to the construction of an affine extractor in Section 10, although their purposes and the techniques used to analyze them are substantially different.

The following theorem and its corollaries are the main results of this section.

Theorem 6.3. *For every $b \in \mathbb{A}^m$ and every affine variety $V \subseteq \mathbb{A}^n$ over $\mathbb{F}$ of dimension at most k and degree at most d, the fiber $(\varphi|_V)^{-1}(b) = \varphi^{-1}(b) \cap V$ is a finite set.*

Corollary 6.4. *φ is an (n, m, k, d) deterministic rank condenser for varieties. In particular, if $m = k$, then φ is an (n, m, d) deterministic rank extractor for varieties.*

We also show in the full version of our paper that the integers $d_1, \ldots, d_n$ and the matrix A can be efficiently constructed.

So we have the following corollary.

Corollary 6.5. *For $m \in \{1, n-1, n\}$, there exists an explicit construction of an (n, m, d) deterministic rank extractor for varieties that is defined by polynomials $f_1, \ldots, f_m \in \mathbb{F}[X_1, \ldots, X_n]$ satisfying the following:*

- *All the coefficients of $f_1, \ldots, f_m$ are in $\{0, 1, -1\}$, and hence are in every subfield of $\mathbb{F}$.*
- *$\deg f_1, \ldots, \deg f_m = O((n+d)\log(n+d))$. And the sparse representations of $f_1, \ldots, f_m$ can be computed in time $\mathrm{poly}(n, d)$. The time complexity can be improved to $\mathrm{poly}(n, \log d)$ at the cost of increasing the degrees of $f_1, \ldots, f_m$ to $O(nd \log n)$.*

[2] We define the minimum distance of the zero code $\{0\}$ to be $n+1$, so that the statement also holds for $k = n$.

[3] While [11] assumes $d_1 > \cdots > d_n$, this assumption does not really matter.

A similar statement holds for general $m \in [n]$ and the coefficients of $f_1, \ldots, f_m$ can be chosen in a finite field $\mathbb{F}_q$, assuming $\mathbb{F}_q$ is a subfield of $\mathbb{F}$ and $q \geq n-1$. The time complexity would also depend polynomially on $\log q$.

The above explicit (n, m, d) deterministic extractor for varieties will be used in the proof of Theorem 1, but only in the case where $m = 1$. Previously, Dvir [9, Theorem 3.1] gave an explicit construction of an $(n, 1, d)$ deterministic rank extractor for varieties, where the polynomial defining the rank extractor is recursively constructed and has degree $\mathrm{poly}(d^n)$. Corollary 6.5 improves the degree of the polynomial to $\widetilde{O}(n+d)$ or $\widetilde{O}(nd)$.

The proof of Theorem 6.3 appears in the full version of this paper.

7 DECOMPOSITION AND MIN-ENTROPY ESTIMATION OF (n, k, d) ALGEBRAIC SOURCES

In this section, we prove that every (n, k, d) algebraic source can be (approximately) decomposed into a convex combination of irreducible, or even irreducibly minimal (n, k, d) sources. In particular, this reduces the problem of constructing deterministic extractors for general (n, k, d) algebraic sources to that for irreducibly minimal (n, k, d) algebraic sources. We will use this reduction in Section 8.

In addition, we show that every (n, k, d) algebraic source D over $\mathbb{F}_q$ is close to a distribution with min-entropy about $k \log q$, and that this estimation is tight up to an additive term of order $O(\log d)$ assuming that k is maximized, i.e., that D is not an $(n, k+1, d)$ algebraic source over $\mathbb{F}_q$.

7.1 Decomposition of (n, k, d) Algebraic Sources

In the full version of the paper, we prove the following lemma:

Lemma 7.1 (Decomposition into irreducible sources). *Suppose $q \geq \max\{20d^5, 2d^2/\varepsilon\}$, where $\varepsilon \in (0,1)$. Then every (n, k, d) algebraic source D over $\mathbb{F}_q$ is ε-close to a convex combination of irreducible (n, k, d) algebraic sources D_i over $\mathbb{F}_q$. Moreover, if D is a minimal (n, k, d) algebraic source over $\mathbb{F}_q$, then each D_i can be chosen to be an irreducibly minimal (n, k, d) algebraic source over $\mathbb{F}_q$.*

Next, we further decompose an irreducible (n, k, d) algebraic source into a convex combination of irreducibly minimal (n, k, d) algebraic sources. Our main tool is the effective fiber dimension theorem (Theorem 4.6). Using this theorem and the results of Section 4, we intersect the variety V with various translates of a carefully chosen linear subspace. There are some bad events that could happen for some of these intersections. For example, the intersection may have the "wrong" dimension, or the resulting variety might have the "correct" dimension k but none of the irreducible components of dimension k are absolutely irreducible. Using the effective fiber dimension theorem, we are able to show that these bad events correspond to small portions of the variety V, and then we again obtain a natural way to decompose the remaining part as a convex combination of irreducibly minimal (n, k, d) sources.

Lemma 7.2. *Suppose $q \geq \max\{20d^5, 2(k+1)d^2/\varepsilon^2\}$, where $\varepsilon \in (0,1)$. Then every irreducible (n, k, d) algebraic source over $\mathbb{F}_q$ is 3ε-close to a convex combination of irreducibly minimal (n, k, d) algebraic sources over $\mathbb{F}_q$.*

Combining Lemma 7.1 and Lemma 7.2 yields the following corollary.

Corollary 7.3 (Decomposition into irreducibly minimal algebraic sources). *Suppose $q \geq \max\{20d^5, 2(k+1)d^2/\varepsilon^2\}$, where $\varepsilon \in (0,1)$. Then every (n, k, d) algebraic source over $\mathbb{F}_q$ is 4ε-close to a convex combination of irreducibly minimal (n, k, d) algebraic sources over $\mathbb{F}_q$.*

7.2 Estimating the Min-Entropy of (n, k, d) Algebraic Sources

We prove the following lower bound on the min-entropy of an (n, k, d) algebraic source D (or more precisely, a distribution D' close to D). The proof uses the decomposition into irreducible (n, k, d) algebraic sources (Lemma 7.1).

Lemma 7.4. *Suppose $q \geq \max\{20d^5, 2kd^2/\varepsilon\}$, where $\varepsilon \in (0, 1/2]$. Then every (n, k, d) algebraic source over $\mathbb{F}_q$ is 2ε-close to a k'-source over the set $\mathbb{F}_q^n$, where $k' = k \log q - \log d - 2$.*

The next proposition complements Lemma 7.4 and gives an upper bound on the min-entropy.

Proposition 7.5. *Suppose $q \geq 20d^5$. Let D be an (n, k, d) algebraic source over $\mathbb{F}_q$ such that k is maximal with respect to this condition, i.e., D is not an $(n, k+1, d)$ algebraic source over $\mathbb{F}_q$. Then the statistical distance between D and any $(k \log q + 2\log d + 2)$-source is at least $\frac{1}{4d}$. Moreover, if D is an irreducible (n, k, d) algebraic source over $\mathbb{F}_q$, then the statistical distance between D and any $(k \log q + \log d + 1)$-source is at least $\frac{1}{2}$.*

8 EXTRACTING A SHORT SEED

In this section, we consider the problem of constructing explicit deterministic extractors for (n, k, d) algebraic sources over a finite field $\mathbb{F}_q$ in the special case where $k = 1$.

The main results of this section are explicit constructions of deterministic extractors that extract almost $\log q$ bits from $(1, 1, d)$ algebraic sources and, more generally, $(n, 1, d)$ algebraic sources over $\mathbb{F}_q$. They are used as building blocks in the construction of the full-fledged deterministic extractors that extract most min-entropy from (n, k, d) algebraic sources.

Formally, we prove the following theorems.

Theorem 8.1 (Extractor for $(1, 1, d)$ algebraic sources). *Let $d \in \mathbb{N}^+$ and $\varepsilon \in (0, 1/2]$. Suppose $q \geq c_0 d^5/\varepsilon^2$, where $c_0 > 0$ is a large enough absolute constant. Then there exists an explicit ε-extractor $\mathsf{Ext} : \mathbb{F}_q \to \{0,1\}^m$ for $(1, 1, d)$ algebraic sources over $\mathbb{F}_q$ such that $m \geq \log q - 2\log\log p - O(\log(d/\varepsilon))$.*

Theorem 8.2 (Extractor for $(n, 1, d)$ algebraic sources). *Let $d \in \mathbb{N}^+$ and $\varepsilon \in (0, 1/2]$. Suppose $q \geq (nd/\varepsilon)^{c_0}$, where $c_0 > 0$ is a large enough absolute constant. Then there exists an explicit ε-extractor $\mathsf{Ext} : \mathbb{F}_q \to \{0,1\}^m$ for $(n, 1, d)$ algebraic sources over $\mathbb{F}_q$ such that $m \geq \log q - 2\log\log p - O(\log(nd/\varepsilon))$.*

Theorem 8.2 is derived from Theorem 8.1. As in [9, 10], the proof of Theorem 8.1 uses Bombieri's estimate for exponential sums (Theorem 4.3). However, the argument in [9, 10] works only when the characterisitic p is large. Moreover, it only yields an extractor that extracts $c \log q$ bits for some constant $c \leq 1/2$. We introduce

new ideas that allow us to extract almost $\log q$ bits regardless of the characteristic p.

As one of our main tools, we prove the following estimate for exponential sums over curves, even over finite fields of small characteristics. Recall that Bombieri's estimate (Theorem 4.3) is valid as long as the polynomial f does not have the form $g^p - g$ on the curve. One way to deal with this difficulty is to require p to be large. However, we would like to get meaningful results for arbitrary p, and we do this by paying the cost of excluding a small subgroup of characters from the estimate.

Lemma 8.3. *Let $C \subseteq \mathbb{A}^n_{\mathbb{F}_q}$ be an irreducible affine curve of degree d_1 over a finite field $\mathbb{F}_q$ of characteristic p, and let $f \in \mathbb{F}_q[X_1, \ldots, X_n]$ be a polynomial of degree d_2 that is not constant on C. Then the set of characters $\chi \in \widehat{\mathbb{F}_q}$ for which*

$$\left| \sum_{x \in C(\mathbb{F}_q)} \chi(f(x)) \right| \leq (d_1^2 + 2d_1d_2 - 3d_1)q^{1/2} + d_1^2$$

fails to hold is contained in a subgroup of $\widehat{\mathbb{F}_q}$ of size at most d_1d_2.

9 DETERMINISTIC EXTRACTORS FOR (n, k, d) ALGEBRAIC SOURCES

In this section, we provide our main construction of deterministic extractors for (n, k, d) algebraic sources. Recall that in Section 8 we considered the case of $(n, 1, d)$ algebraic sources.

We start with the case of (n, n, d) algebraic sources, and we follow our general proof technique as laid out in Section 1.3: the first step of the construction is applying our extractor from Section 8 to obtain a short output, which is then, in the second step, used as a seed for a seeded extractor for sources with high min-entropy (note that even though we have more structure in our source, since we are anyway applying a seeded extractor we might as well use an off-the-shelf construction which works for any source with high min-entropy). Proving that this indeed works requires analyzing the conditional distribution of an (n, n, d) algebraic source under fixing of a subset of the coordinates, which is done in the full version of the paper. This construction is presented and analyzed in Section 9.1.

In order to remove the assumption that $k = n$ and handle general (n, k, d) algebraic sources, we apply a rank extractor which, roughly speaking, condenses a k-dimensional source in an ambient n-dimensional space to a k-dimensional source in an ambient k-dimensional space, and this enables us to use the extractor from Section 9.1. As discussed at the end of Section 9.1, this can be done using the deterministic rank extractor of Section 6, but it would have an undesirable effect on the field size. Thus, we opt to use a *linear seeded rank extractor* (as defined in Section 5), where the seed of the rank extractor is chosen pseudorandomly using our extractor for $(n, 1, d)$ algebraic sources from Section 8.

To summarize, in our composition theorem (Theorem 9.3), we start by applying the extractor for $(n, 1, d)$ algebraic sources from Section 8 in order to select a seed for the seeded linear rank extractor from Section 5, we apply the resulting linear map to the source, and then we use the extractor for full-rank sources from Section 9.1 to obtain the final output. The details of this construction appear in Section 9.2.

9.1 Deterministic Extractors for Full-Rank Algebraic Sources

We need the following explicit construction of seeded extractors given by Goldreich and Wigderson [17], which is based on expander graphs.

Theorem 9.1 ([17]). *For $n \in \mathbb{N}$, $0 \leq \Delta \leq n$ and $\varepsilon > 0$, there exists an explicit seeded ε-extractor $\mathsf{Ext} : \{0,1\}^n \times \{0,1\}^\ell \to \{0,1\}^n$ for $(n - \Delta)$-sources with $\ell = O(\Delta + \log(1/\varepsilon))$.*

We now state our construction for full-rank algebraic sources. Our construction follows the general paradigm mentioned in Section 1.3: we first apply our extractor from Theorem 8.1 to obtain a short output, which is then used as a seed to the extractor from Theorem 9.1.

Theorem 9.2 (Extractor for (n, n, d) algebraic sources). *Let $n, d \in \mathbb{N}^+$ and $\varepsilon \in (0, 1/2]$. Suppose $q \geq (nd/\varepsilon)^{c_0}$, where $c_0 > 0$ is a large enough absolute constant. Then there exists an explicit ε-extractor $\mathsf{Ext} : \mathbb{F}_q \to \{0,1\}^m$ for (n, n, d) algebraic sources over $\mathbb{F}_q$ such that $m \geq n \log q - 2 \log\log p - O(\log(d/\varepsilon))$.*

One can remove the full-rank assumption and construct an extractor for (n, k, d) algebraic sources over $\mathbb{F}_q$ by composing the extractor in Theorem 9.2 with the deterministic rank extractor for varieties in Section 6. This argument was used by Dvir, Gabizon and Wigderson [10], except that they considered polynomial sources only and used a different construction of deterministic rank extractors. The downside of this argument, however, is that such a deterministic rank extractor is necessarily nonlinear. In particular, our rank extractor uses polynomials of degree at least $\mathrm{poly}(n)$, and so does the one in [10]. Composing with such a rank extractor increases the degree of each polynomial in the polynomial map by at least a $\mathrm{poly}(n)$ factor. The resulting field size q would then depend at least polynomially on n^k, or n^n if $k = \Theta(n)$, assuming that we want to extract about $k \log q$ bits.

In the next subsection, we show how to remove the full-rank assumption more efficiently using a linear seeded rank extractor for varieties.

9.2 Removing the Full-Rank Assumption

We now remove the full-rank assumption in Theorem 9.2 without significantly increasing the required field size. This is done by extending an argument in [15, 16].

The following theorem shows how to compose all the ingredients in our construction: an extractor Ext_1 for $(n, 1, d)$ algebraic sources, an extractor Ext_2 for full-rank algebraic sources, and a linear seeded rank extractor φ, in order to obtain extractors for (n, k, d) algebraic sources. The construction uses Ext_1 in order to select the seed for φ, applies φ on the input, and then applies Ext_2 on the resulting "condensed" source.

Theorem 9.3 (Composition of extractors). *Let $n \geq k > 1$ be integers. Let $\varepsilon, \varepsilon' \in (0, 1)$. Suppose we are given the following objects:*

- *an ε-extractor $\mathsf{Ext}_1 : \mathbb{F}_q^n \to \{0,1\}^{m_1}$ for $(n, 1, d)$ algebraic sources over $\mathbb{F}_q$,*
- *an ε-extractor $\mathsf{Ext}_2 : \mathbb{F}_q^{k-1} \to \{0,1\}^{m_2}$ for $(k-1, k-1, d)$ algebraic sources over $\mathbb{F}_q$, and*

- *an $(n, k-1, k, \varepsilon')$ linear seeded rank extractor $(\varphi_y)_{y\in\{0,1\}^\ell}$ for varieties over $\overline{\mathbb{F}}_q$ (see Definition 5.1) such that $\ell \le m_1$ and each φ_y is defined by linear polynomials over $\mathbb{F}_q$.*

Write $\mathsf{Ext}_1 = (\mathsf{Ext}_1', \mathsf{Ext}_1'')$, where Ext_1' and Ext_1'' output the first ℓ bits and the last $m_1 - \ell$ bits of Ext_1 respectively. Assume $q \ge \max\{20d^5, 2(k+1)d^2/\varepsilon^2\}$. Then the map $\mathsf{Ext} : \mathbb{F}_q^n \to \{0,1\}^{m_1} \times \{0,1\}^{m_2} = \{0,1\}^{m_1+m_2}$ defined by

$$\mathsf{Ext}(x) := (\mathsf{Ext}_1(x), \mathsf{Ext}_2(\varphi_{\mathsf{Ext}_1'(x)}(x)))$$

is a $(6\varepsilon \cdot 2^\ell + 4\varepsilon + \varepsilon')$-extractor for (n, k, d) algebraic sources over $\mathbb{F}_q$.

Instantiating the objects in Theorem 9.3 immediately implies Theorem 1. The details appear in the full version of this paper.

10 AFFINE EXTRACTORS WITH EXPONENTIALLY SMALL ERROR FOR QUASIPOLYNOMIALLY LARGE FIELDS

In this section, we construct affine extractors with exponentially small error, over prime fields of size $q = n^{O(\log\log(n))}$ and any characteristic. Our construction is in fact identical to the extractor of Bourgain, Dvir and Leeman [5], but our analysis is slightly improved. Specifically, Bourgain, Dvir and Leeman constructed an affine extractor over prime fields $\mathbb{F}_q$ where $q = n^{O(\log\log n)}$ is a so-called "typical" prime. Our construction works over any prime finite field of the same size.

The following proposition replaces the use of [5] by finding a set of degrees $d_1, \ldots, d_n$ with useful properties for the construction.

Proposition 10.1. *Let q be a prime number. Fix $\varepsilon > 0$. Then, if $q \ge n^{\frac{2}{\varepsilon}\log\log(n)}$, there exists an efficient deterministic algorithm that, in time polynomial in n, finds n integers $d_1 < d_2 < \cdots < d_n \in \mathbb{N}$ such that $LCM(d_1, \ldots, d_n) \le q^\varepsilon$ and each d_i is coprime to $q-1$.*

Let $A \in \mathbb{F}^{m\times n}$ be a matrix where every m columns are linearly independent (e.g., a Vandermonde matrix). Let $d_1 < d_2 < \cdots < d_n$ be as in Proposition 10.1 and define the function $E : \mathbb{F}^n \to \mathbb{F}^m$ by

$$E(x_1, \ldots, x_n) = A \cdot \begin{pmatrix} x_1^{d_1} \\ \vdots \\ x_n^{d_n} \end{pmatrix}. \tag{1}$$

Theorem 10.2. *For every $0 < \beta < 1/2$, there exists a constant C such that the following holds: Let $k \le n$ be integers and $\mathbb{F}$ be a prime field of size $q \ge n^{C\log\log n}$. Then for $m = \beta k$ the function $E : \mathbb{F}^n \to \mathbb{F}^m$ as in (1) is an affine extractor for min-entropy k with error $q^{-\Omega(k)}$. That is, for every affine subspace $V \subseteq \mathbb{F}^n$ of dimension k, if X_V is a random variable uniformly distributed on V, $E(X_V)$ is $q^{-\Omega(k)}$-close to uniform on $\mathbb{F}^k$.*

11 EXPLICIT NOETHER NORMALIZATION FOR AFFINE VARIETIES AND AFFINE ALGEBRAS

The Noether normalization lemma [26, 28] is a cornerstone of commutative algebra and algebraic geometry. It states that any finitely generated commutative algebra over a field $\mathbb{F}$, or what we call an *affine algebra* over $\mathbb{F}$, is not too far from a polynomial ring, in the sense that it is always a finitely generated module over a subring that is isomorphic to a polynomial ring $\mathbb{F}[Y_1, \ldots, Y_k]$. The geometric interpretation of this statement is that any affine variety V over $\mathbb{F}$ is a "branched covering" of an affine space $\mathbb{A}^k_{\mathbb{F}}$, or more precisely, V admits a surjective finite morphism $\varphi_V : V \to \mathbb{A}^k_{\mathbb{F}}$.

When $\mathbb{F}$ is an infinite field (or more generally, a sufficiently large field), the polynomials that define the finite morphism φ_V may be chosen to be linear polynomials (see, e.g., Lemma 4.4). In general, φ_V can always be chosen to be defined by polynomials of sufficiently large degrees. In fact, counting arguments show that given the variety, a "random" polynomial map defined by polynomials of sufficiently large degrees would almost surely yield such a finite morphism. See [6] for a quantitative analysis. However, it is not known how to completely "derandomize" such counting arguments.

The first proof of the Noether normalization lemma for general affine algebras over arbitrary fields was given by Nagata [24–26]. This proof has the interesting feature that it actually constructs a "universal" polynomial map $\varphi : \mathbb{A}^n_{\mathbb{F}} \to \mathbb{A}^k_{\mathbb{F}}$ that works for all low-degree affine varieties. Namely, for any low-degree affine variety $V \subseteq \mathbb{A}^n_{\mathbb{F}}$ of dimension k, the restriction of φ to V gives a finite morphism $\varphi|_V : V \to \mathbb{A}^k_{\mathbb{F}}$. The existence of such a polynomial map φ that is independent of V appears to be stronger and more intriguing than the existence of finite morphisms $V \to \mathbb{A}^k_{\mathbb{F}}$. In fact, we do not know how to prove the existence of φ via a counting argument.

While the polynomial map φ constructed by Nagata gives a uniform way of constructing finite morphisms, a drawback is that the degrees of the polynomials that define φ can get extremely high due to the iterative nature of the construction. More specifically, the map φ is constructed as a composition of polynomial maps $\varphi_i : \mathbb{A}^{i+1}_{\mathbb{F}} \to \mathbb{A}^i_{\mathbb{F}}$, $i = n-1, \ldots, k$ such that their restrictions $\varphi_i|_{V_{i+1}}$ are finite morphisms, where we inductively define $V_n = V$ and $V_i = \overline{\varphi_i(V_{i+1})}$ for $i = n-1, \ldots, k$. The problem is that composing with a polynomial map can increase the degree of a variety exponentially (see Lemma 4.8). The degree bound for the polynomials defining φ is at least doubly exponential for this reason.

Thus, it is a natural question to ask if there is a more efficient construction of the universal polynomial map φ. In this section, we show that the DKL construction in Section 6 is indeed such a construction, which always works when $|\mathbb{F}| \ge n$.

The construction of φ. We first recall the DKL construction in Section 6. Let $\mathbb{F}$ be a field. Let $n, d \in \mathbb{N}^+$ and $m, k \in [n]$. Let $d_1, \ldots, d_n$ be n pairwise coprime integers greater than d. Let $M = (c_{i,j})_{i\in[m], j\in[n]} \in \mathbb{F}^{m\times n}$ be a k-regular matrix, i.e., any k distinct columns of M are linearly independent. Let $\varphi = \varphi(M) : \mathbb{A}^n_{\mathbb{F}} \to \mathbb{A}^m_{\mathbb{F}}$ be the polynomial map defined by $f_1, \ldots, f_m \in \mathbb{F}[X_1, \ldots, X_n]$, where $f_i := \sum_{j=1}^n c_{i,j} X_j^{d_j}$. In other words, φ is given by

$$\varphi : (a_1, \ldots, a_n) \mapsto \left(\sum_{j=1}^n c_{1,j} a_j^{d_j}, \ldots, \sum_{j=1}^n c_{m,j} a_j^{d_j} \right).$$

The main results of this subsection are the following theorems.

Theorem 11.1 (Explicit Noether normalization for affine varieties). *Let V be an affine variety of dimension at most k and degree at most d over a field $\mathbb{F}$. Then $\varphi|_V : V \to \mathbb{A}^m_{\mathbb{F}}$ is a finite morphism.*

Theorem 11.1 translates into the following algebraic statement, Theorem 11.2, which gives an explicit Noether normalization lemma for affine algebras, i.e., finitely generated commutative algebras over a field.

Recall that the *Krull dimension* of a commutative ring A is the supremum of the lengths of all chains of prime ideals in A. If V is an affine variety over a field $\mathbb{F}$, then the Krull dimension of its coordinate ring $\mathbb{F}[V]$ is just the dimension of V.

Theorem 11.2 (Explicit Noether normalization for affine algebras). *Suppose A is a commutative $\mathbb{F}$-algebra generated by $a_1, \ldots, a_n \in A$ such that the Krull dimension of A is at most k. Let the ideal I of $\mathbb{F}[X_1, \ldots, X_n]$ be the ideal of all polynomial relations satisfied by $a_1, \ldots, a_n$. Also suppose the degree of the affine variety $V(I) \subseteq \mathbb{A}^n$ is at most d. Then A is a finitely generated module over its subring $S = \mathbb{F}[f_1(a), \ldots, f_m(a)]$, where $f_1, \ldots, f_m$ are the polynomials defining φ and $a = (a_1, \ldots, a_n)$.*

The fact that A is a finitely generated module over S implies that the Krull dimension of S equals that of A. In the case where the Krull dimension of A is k and $k = m$, this means $f_1(a), \ldots, f_m(a)$ are algebraically independent over $\mathbb{F}$, and hence S is isomorphic to a polynomial ring $\mathbb{F}[Y_1, \ldots, Y_m]$ via $f_i(a) \mapsto Y_i$.

Theorem 11.1 and Theorem 11.2 are proved in the full version of this paper. The proof is inspired by and closely follows a geometric proof sketched in [20, Remark 1].

Smaller fields. While $k \times n$ MDS matrices are generally not known over small finite fields $\mathbb{F}_q$, which prevents us from choosing $m = k$ over $\mathbb{F}_q$, it may still be possible to choose larger m for which (explicit) k-regular $m \times n$ matrices over $\mathbb{F}_q$ exist, and this would yield a finite morphism $\varphi|_V : V \to \mathbb{A}^m_{\mathbb{F}_q}$ by Theorem 11.1. As compositions of finite morphisms are finite [2, Corollary 5.4], by replacing n with m and V with $V' = \overline{\varphi(V)}$, we reduce the problem of constructing a finite morphism on $V \subseteq \mathbb{A}^n_{\mathbb{F}_q}$ to constructing that on $V' \subseteq \mathbb{A}^m_{\mathbb{F}_q}$, where V' has the same dimension as V but lives in a possibly much smaller affine space $\mathbb{A}^m_{\mathbb{F}_q}$. The degree of V', however, may be significantly larger than that of V. See Lemma 4.8 for a general upper bound on the degree.

For example, while we do not know the existence of $k \times n$ MDS matrices over small finite fields $\mathbb{F}_q$, one can still use a BCH-code-like construction to obtain an $m \times n$ k-regular matrix with $m = O(k \log_q n)$, which can be much smaller than n if $k \ll n$. Applying the resulting map φ reduces the dimension of the ambient space from n to m.

However, when q is really small and k is close to n, it may be possible that one can only choose $m = n - 1$ and hence only reduce the dimension of the ambient space by one at each step. This is essentially the same method used in Nagata's construction. Currently, all constructions of the universal polynomial map $\varphi : \mathbb{A}^n_{\mathbb{F}_q} \to \mathbb{A}^k_{\mathbb{F}_q}$ with $k = \dim V$ that we know over a constant-size field $\mathbb{F}_q$ use polynomials of degree at least doubly exponential in $\min\{k, n - k\}$ due to the blow-up of the degree of the variety. It is an interesting question to ask if there exist constructions with a better degree bound over constant-size fields.

ACKNOWLEDGMENTS

Zeyu Guo was supported in part by a Simons Investigator Award (#409864, David Zuckerman). David Zuckerman was supported in part by NSF Grants CCF-1705028 and CCF-2008076, a Simons Investigator Award (#409864), and the Center of Mathematical Sciences and Applications at Harvard University.

REFERENCES

[1] Divesh Aggarwal, Huck Bennett, Alexander Golovnev, and Noah Stephens-Davidowitz. 2021. Fine-grained hardness of CVP(P) - Everything that we can prove (and nothing else). In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021*. SIAM, 1816–1835. https://doi.org/10.1137/1.9781611976465.109

[2] Michael F. Atiyah and I. G. MacDonald. 1969. *Introduction to Commutative Algebra*. Addison-Wesley-Longman.

[3] Enrico Bombieri. 1966. On exponential sums in finite fields. *Amer. J. Math.* 88 (1966), 71–105. https://doi.org/10.2307/2373048

[4] Jean Bourgain. 2007. On the construction of affine extractors. *GAFA Geometric And Functional Analysis* 17, 1 (2007), 33–57. https://doi.org/10.1007/s00039-007-0593-z

[5] Jean Bourgain, Zeev Dvir, and Ethan Leeman. 2016. Affine extractors over large fields with exponential error. *computational complexity* 25, 4 (2016), 921–931. https://doi.org/10.1007/s00037-015-0108-5

[6] Juliette Bruce and Daniel Erman. 2019. A probabilistic approach to systems of parameters and Noether normalization. *Algebra & Number Theory* 13, 9 (2019), 2081–2102. https://doi.org/10.2140/ant.2019.13.2081

[7] Eshan Chattopadhyay, Jesse Goodman, and Jyun-Jie Liao. 2021. Affine extractors for almost logarithmic entropy. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. https://doi.org/10.1109/FOCS52979.2021.00067

[8] Eshan Chattopadhyay and David Zuckerman. 2019. Explicit Two-Source Extractors and Resilient Functions. *Annals of Mathematics* 189 (2019), 653–705. https://doi.org/10.4007/annals.2019.189.3.1

[9] Zeev Dvir. 2012. Extractors for varieties. *computational complexity* 21, 4 (2012), 515–572. https://doi.org/10.1007/s00037-011-0023-3

[10] Zeev Dvir, Ariel Gabizon, and Avi Wigderson. 2009. Extractors and rank extractors for polynomial sources. *computational complexity* 18, 1 (2009), 1–58. https://doi.org/10.1007/s00037-009-0258-4

[11] Zeev Dvir, János Kollár, and Shachar Lovett. 2014. Variety evasive sets. *computational complexity* 23, 4 (2014), 509–529. https://doi.org/10.1007/s00037-013-0073-9

[12] Michael A. Forbes. 2014. *Polynomial identity testing of read-once oblivious algebraic branching programs*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[13] Michael A. Forbes and Venkatesan Guruswami. 2015. Dimension Expanders via Rank Condensers. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015 (LIPIcs, Vol. 40)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 800–814. https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2015.800

[14] Michael A. Forbes and Amir Shpilka. 2012. On identity testing of tensors, low-rank recovery and compressed sensing. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*. 163–172.

[15] Ariel Gabizon and Ran Raz. 2008. Deterministic extractors for affine sources over large fields. *Combinatorica* 28, 4 (2008), 415–440. https://doi.org/10.1007/s00493-008-2259-3

[16] Ariel Gabizon, Ran Raz, and Ronen Shaltiel. 2006. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SIAM J. Comput.* 36, 4 (2006), 1072–1094. https://doi.org/10.1137/S0097539705447049

[17] Oded Goldreich and Avi Wigderson. 1997. Tiny families of functions with random properties: A quality-size trade-off for hashing. *Random Structures & Algorithms* 11, 4 (1997), 315–343. https://doi.org/10.1002/(SICI)1098-2418(199712)11:4<315::AID-RSA3>3.0.CO;2-1

[18] Alexander Golovnev, Alexander S. Kulikov, and R. Ryan Williams. 2021. Circuit Depth Reductions. In *12th Innovations in Theoretical Computer Science Conference, ITCS 2021 (LIPIcs, Vol. 185)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 24:1–24:20. https://doi.org/10.4230/LIPIcs.ITCS.2021.24

[19] Pavel Hrubeš and Anup Rao. 2015. Circuits with Medium Fan-In. In *30th Conference on Computational Complexity, CCC 2015, June 17-19, 2015, Portland, Oregon, USA (LIPIcs, Vol. 33)*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 381–391. https://doi.org/10.4230/LIPIcs.CCC.2015.381

[20] János Kollár, Lajos Rónyai, and Tibor Szabó. 1996. Norm-graphs and bipartite Turán numbers. *Combinatorica* 16, 3 (1996), 399–406. https://doi.org/10.1007/BF01261323

[21] Fu Li and David Zuckerman. 2019. Improved extractors for recognizable and algebraic sources. In *23rd International Conference on Randomization and Computation*

(RANDOM). https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2019.72

[22] Xin Li. 2011. A New Approach to Affine Extractors and Dispersers. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity, CCC 2011, San Jose, California, USA, June 8-10, 2011*. IEEE Computer Society, 137–147. https://doi.org/10.1109/CCC.2011.27

[23] Xin Li. 2016. Improved Two-Source Extractors, and Affine Extractors for Polylogarithmic Entropy. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016*. IEEE Computer Society, 168–177. https://doi.org/10.1109/FOCS.2016.26

[24] Masayoshi Nagata. 1953. Some remarks on local rings. *Nagoya Mathematical Journal* 6 (1953), 53–58. https://doi.org/10.1017/S0027763000016974

[25] Masayoshi Nagata. 1956. A general theory of algebraic geometry over Dedekind domains, I: the notion of models. *American Journal of Mathematics* 78, 1 (1956), 78–116. https://doi.org/10.2307/2372486

[26] Masayoshi Nagata. 1962. *Local Rings*. New York, Interscience Publishers.

[27] Joseph Naor and Moni Naor. 1993. Small-bias probability spaces: efficient constructions and applications. *SIAM J. Comput.* 22, 4 (1993), 838–856. https://doi.org/10.1137/0222053

[28] Emmy Noether. 1926. Der Endlichkeitssatz der Invarianten endlicher linearer Gruppen der Charakteristik p. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1926 (1926), 28–35.

[29] Anup Rao. 2007. An Exposition of Bourgain's 2-Source Extractor. In *TR 07-034*. Electronic Colloqium on Computaitonal Complexity.

[30] Zachary Remscrim. 2016. The Hilbert function, algebraic extractors, and recursive Fourier sampling. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 197–208. https://doi.org/10.1145/2213977.2213995

[31] Igor R. Shafarevich. 1994. *Basic Algebraic Geometry 1: Varieties in Projective Space*. Springer-Verlag. https://doi.org/10.1007/978-3-642-37956-7

[32] Salil Vadhan. 2012. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science* 7, 1-3 (2012), 1–336. https://doi.org/10.1561/0400000010

[33] Ravi Vakil. 2022. The Rising Sea: Foundations of Algebraic Geometry. https://math.stanford.edu/~vakil/216blog/FOAGaug2922publici.pdf. August 29, 2022 version.

[34] Amir Yehudayoff. 2011. Affine extractors over prime fields. *Combinatorica* 31, 2 (2011), 245–256. https://doi.org/10.1007/s00493-011-2604-9

Received 2022-11-07; accepted 2023-02-06