# Noisy Interpolation of Sparse Polynomials, and Applications

Shubhangi Saraf
MIT
shibs@mit.edu

Sergey Yekhanin
Microsoft Research
yekhanin@microsoft.com

## Abstract

*Let $f \in \mathbb{F}_q[x]$ be a polynomial of degree $d \leq q/2$. It is well-known that $f$ can be uniquely recovered from its values at some $2d$ points even after some small fraction of the values are corrupted. In this paper we establish a similar result for sparse polynomials. We show that a $k$-sparse polynomial $f \in \mathbb{F}_q[x]$ of degree $d \leq q/2$ can be recovered from its values at $O(k)$ randomly chosen points, even if a small fraction of the values of $f$ are adversarially corrupted.*

*Our proof relies on an iterative technique for analyzing the rank of a random minor of a matrix. We use the same technique to establish a collection of other results. Specifically,*

- *We show that restricting any linear $[n, k, \delta n]_q$ code to a randomly chosen set of $O(k)$ coordinates with high probability yields an asymptotically good code.*
- *We improve the state of the art in locally decodable codes, showing that similarly to Reed Muller codes matching vector codes require only a constant increase in query complexity in order to tolerate a constant fraction of errors. This result yields a moderate reduction in the query complexity of the currently best known codes.*
- *We improve the state of the art in constructions of explicit rigid matrices. For any prime power $q$ and integers $n$ and $d$ we construct an explicit matrix $M$ with $\exp(d) \cdot n$ rows and $n$ columns such that the rank of $M$ stays above $n/2$ even if every row of $M$ is arbitrarily altered in up to $d$ coordinates. Earlier, such constructions were available only for $q = O(1)$ or $q = \Omega(n)$.*

## 1 Introduction

Sparse polynomial interpolation over finite fields when input data are exact has numerous applications and has been studied by many authors [7, 12, 22, 19, 31]. There is also a broad literature addressing the problem of noisy (non-sparse) polynomial interpolation. In fact, much of the research in algebraic coding theory can be seen as dealing with precisely this problem, e.g., [2, 14].

The intersection of the two problems above, i.e., the problem of noisy interpolation of sparse polynomials has also attracted some attention in recent years, e.g., [26, 27] give a probabilistic polynomial time algorithm to recover a $k$-sparse univariate polynomial $f \in \mathbb{F}_p[x]$ with a known set of monomial degrees and total degree below (roughly) $O(p/\log p)$ from the noisy values at $\mathrm{poly(k, \log p)}$ randomly chosen points, where the noise is bounded in the Lee metric. A multiplicative analog of the above problem over the integers has been addressed in [30]. Yet,

some very basic questions about noisy interpolation of sparse polynomials have remained open. In this paper we deal with one of such questions.

Let $L$ be the collection of all $k$-sparse polynomials (with some fixed set of monomial degrees) in $\mathbb{F}_q[x]$. Assume that the maximal degree is below $(1-\delta)q$. We say that a set $S \subseteq \mathbb{F}_q$ is an $\alpha$-noisy interpolating set for $L$, if any polynomial in $L$ can be uniquely recovered from its values on the set $S$, even after at most $\alpha|S|$ of these values have been corrupted arbitrarily.

We study the size of the smallest noisy interpolating set for $k$-sparse polynomials. The simple lower bound for this quantity is $\Omega(k)$; the simple upper bound (following from the Chernoff bound) is $O_{\delta,\alpha}(k \log q)$. We show that the lower bound is essentially tight, i.e., that a random subset of $\mathbb{F}_q$ of size $O_\delta(k)$ suffices for $\delta/16$-noisy sparse interpolation. The key feature of our result is that the bound we get is independent of the field size.

Our proof uses an iterative argument to analyze the rank of a random minor of a matrix. The argument generalizes to show that a random restriction of any $[n, k, \delta n]_q$ linear code to a set of $O(k)$ coordinates is asymptotically good with high probability. Our results find applications to locally decodable codes and our technique gives improved constructions of explicit rigid matrices. We elaborate on these applications below.

### 1.1 Locally decodable codes

Locally Decodable Codes (LDCs) are error-correcting codes that admit highly efficient sub-linear time decoding algorithms. An $r$-query locally decodable code $C$ encodes $k$-symbol messages $\mathbf{x}$ to $N$-symbol codewords $C(\mathbf{x})$ in such a way that one can probabilistically recover any symbol $\mathbf{x}(i)$ of the message by querying only $r \ll k$ symbols of the (possibly corrupted) codeword. Ideally, one would like to have codes where both the codeword length $N(k)$ and the query complexity $r(k)$ are as low as possible. It turns out however that one cannot minimize both of these parameters simultaneously. There is a trade-off. Understanding the true shape of the trade-off, and constructing optimal codes are the key goals of the LDC related research [33].

Early constructions of LDCs [3] are based on the classical Reed Muller (RM) codes. The code consists of complete evaluations of polynomials of total degree up to $d$ in $\mathbb{F}_q[z_1, \ldots, z_n]$. The simplest decoder recovers the value of the unknown polynomial $f$ at a point $\mathbf{w}$ by shooting a line $L$ in a random direction through $\mathbf{w}$, querying some $d+1$ points on $L$, and using polynomial interpolation to recover the restriction of $f$ to $L$. This construction yields $(d+1)$-query codes of length $\exp\left(k^{1/d}\right)$ tolerating some $O(1/d)$ fraction of errors.

Note that in the solution above as the number of queries increases, the codeword length becomes smaller as a function of $k$, but at the price of a reduction in the error-rate that the code could handle. This weakness however is easy to overcome [4, 18, 28]. All we need to do is make the decoder issue (say) $2d$ queries instead of $(d+1)$, and use *noisy* polynomial interpolation instead of ordinary polynomial interpolation. Altogether this yields $2d$-query codes of the same length tolerating a fixed constant fraction of errors independent of $d$.

Recently, in a sequence of works [32, 24, 10, 15, 23] a new family of locally decodable codes called Matching Vector (MV) codes has been developed. Codes in this family have dramatically better parameters than Reed Muller based LDCs. Specifically for every integer $t \geq 2$, [10, 15, 23] construct $r = O(2^t)$-query codes that encode $k$-long messages to $\exp\exp\left((\log k)^{1/t}(\log\log k)^{1-1/t}\right)$-long codewords and tolerate some $O(1/r)$ fraction of errors.

Again, we see the issue similar to the one we discussed above arise: as the number of queries increases, the codeword length becomes smaller but the error-rate that the code can handle suffers. This issue has been addressed in [9, 5] where it was shown that increasing the query complexity from $O\left(2^t\right)$ to $t^{O(t)}$ allows one to handle an error rate independent of $t$, and later in [6] where it was shown that just $O\left(t \cdot 2^t\right)$ queries suffice. In this paper we bring this line of work to an end showing that similarly to Reed Muller codes matching vector codes require only a constant multiplicative increase in the query complexity in order to tolerate a fixed constant fraction of errors.[1]

[1] We elaborate a little more on the relation between our results and those of [6]. Our results apply to standard non-binary matching vector codes as given in [10, 9, 5]. The results of [6] however apply to a different (though closely related) family of binary codes. Our

The proof heavily relies on the polynomial view of matching vector codes developed in [9], and on our result on noisy interpolation of sparse polynomials.

### 1.2 Matrix rigidity

The notion of matrix rigidity was introduced by Leslie Valiant in 1977 [29]. Valiant called an $m \times n$ matrix $A$ defined over a field $(r, d)$-rigid, if it is not possible to reduce the rank of $A$ below $r$ by arbitrarily altering each row of $A$ in up to $d$ coordinates. Valiant showed that if a matrix $A \in \mathbb{F}^{m \times n}$ is $(\Omega(n), n^{\epsilon})$-rigid and $m = O(n)$; then the linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ induced by $A$ cannot be computed by a linear circuit that simultaneously has size $O(n)$ and depth $O(\log n)$. Valiant's work has naturally led to the challenge of constructing explicit rigid matrices, since any such matrix yields an explicit linear map, for which we get a circuit lower bound. After more than three decades of efforts, however, this challenge remains elusive [21].

None of the existing techniques for constructing rigid matrices surpasses the basic *untouched minor* argument of [25] that amounts to taking a matrix where every minor has full rank, and using the bound from the unbalanced Zarankiewicz problem [16, p. 29] to show that after up to $d$ arbitrary changes per row there remains a somewhat large minor that has not been touched. Note that the untouched minor argument requires $q = \Omega(n)$ in order to have an explicit matrix with full rank minors.

The definition of an $n$-column rigid matrix involves three parameters. The number or rows $m$, the remaining rank $r$, and the number of allowed changes in a single row $d$. The trade-off between $r$ and $d$ (for $m = n$) has been addressed in [11, 17, 20, 25]. In particular Friedman [11] designed a family of explicit matrices over constant sized fields that meets the parameters coming from the untouched minor argument.[2]

The trade-off between $m$ and $d$ (for $r = n/2$) has been addressed in [1, 8]. In particular Alon et al. [1] designed explicit $(n/2, d)$-rigid matrices over constant sized fields with $m = \exp(d) \cdot n$ meeting the parameters that can be derived over $\mathbb{F}_q$, $q = \Omega(n)$ via the untouched minor technique.

Thus prior to our work, constructions of $(n/2, d)$-rigid matrices with $m = \exp(d) \cdot n$ were available over fields $\mathbb{F}_q$ for $q = O(1)$ and $q = \Omega(n)$. In this paper we close the gap and obtain such constructions for all values of $q$. Our proof largely follows that of [1], with one important new ingredient. Specifically we show that for any linear space $L \subseteq \mathbb{F}_q^n$, $\dim L \leq (1 - \epsilon)n$ there exists a point $\mathbf{x} \in \{0, 1\}^n$ such that $\mathbf{x}$ is $\Omega(n)$-far from $L$.

### 1.3 Organization

In Section 3 we establish our main result, showing that with high probability a random restriction of an arbitrary low rate $q$-ary linear code of good distance is asymptotically good. This entails our result on noisy interpolation of sparse polynomials. In Section 4 we obtain our results on locally decodable codes, and in Section 5 we obtain our results on matrix rigidity.

## 2 Notation

We use the following standard mathematical notation:

- $d(\mathbf{x}, \mathbf{y})$ denotes the Hamming distance between vectors $\mathbf{x}, \mathbf{y}$;
- For a vector $\mathbf{y} \in \mathbb{F}_q^n$, and a set $L \subseteq \mathbb{F}_q^n$, $d(L, \mathbf{y})$ denotes $\min_{\mathbf{x} \in L} d(\mathbf{x}, \mathbf{y})$;

codes have better query complexity and tolerate a fixed fraction of errors. Codes of [6] however have a moderately larger query complexity but tolerate the optimal, i.e., the largest possible fraction of errors (in the list-decoding model).

[2] Interestingly, Friedman's paper [11] actually predates the work of Shokrollahi et al. [25].

- Let $L \subseteq \mathbb{F}_q^n$ be a linear space, and $S \subseteq [n]$ be a multi-set; then $L|_S \subseteq \mathbb{F}_q^{|S|}$ denotes the restriction of $L$ to coordinates in $S$;
- Similarly, let $G \in \mathbb{F}_q^{k \times n}$ be a matrix, and $S \subseteq [n]$ a multi-set; then $G|_S \in \mathbb{F}_q^{k \times |S|}$ denotes the restriction of $G$ to columns whose indices belong to $S$;
- Finally, a linear $[n, k, d]_q$ code is a $k$-dimensional linear subspace of $\mathbb{F}_q^n$, where the minimal Hamming weight of any nonzero vector in the subspace is at least $d$.

## 3 Random restrictions of linear codes with good distance are asymptotically good

In this section we establish our main result (Theorem 3). We start with the following lemmas.

**Lemma 1** *Let $q$ be a prime power, $n$ and $k$ be integers and $\delta$ be positive. Let $G \in \mathbb{F}_q^{k \times n}$ be a generator matrix of a linear $[n, k, \delta n]_q$ code. Suppose a multi-set $S \subseteq [n]$ is such that* $\mathrm{rk}\left(\mathrm{G}|_{\mathrm{S}}\right) < \mathrm{k}$; *then*

$$\mathrm{Pr}_{\mathrm{i} \in [\mathrm{n}]}\left[\mathrm{rk}\left(\mathrm{G}|_{\mathrm{S} \cup \{\mathrm{i}\}}\right) > \mathrm{rk}\left(\mathrm{G}|_{\mathrm{S}}\right)\right] \geq \delta. \tag{1}$$

**Proof:** Let $K \subseteq \mathbb{F}_q^k$ be the kernel of $G|_S$. Clearly, $\dim(K) \geq 1$. We now show that with probability at least $\delta$, $\mathrm{Ker}\left(\mathrm{G}|_{\mathrm{S} \cup \{\mathrm{i}\}}\right) \subset \mathrm{Ker}\left(\mathrm{G}|_{\mathrm{S}}\right)$, which immediately implies (1).

Let $\mathbf{x}$ be any nonzero element in $\mathrm{Ker}\left(\mathrm{G}|_{\mathrm{S}}\right)$. The code generated by $G$ has distance $\delta$, therefore at least $\delta$ fraction of the coordinates of $\mathbf{x}G$ must be nonzero. Thus with probability at least $\delta$, a randomly chosen index $i$ will be such that $(\mathbf{x}G)_i \neq 0$. If so, then clearly $\mathbf{x}$ will not be in the kernel of $G|_{S \cup \{i\}}$. ■

**Lemma 2** *Let $q$ be a prime power, $s, k \leq n$ be integers, and $\delta$ be positive. Let $G$ be a generator matrix of a linear $[n, k, \delta n]_q$ code; then for a random choice of a multi-set $S$ we have*

$$\mathrm{Pr}_{\mathrm{S} \subseteq [\mathrm{n}] \mid |\mathrm{S}| = \mathrm{s}}\left[\mathrm{rk}\left(\mathrm{G}|_{\mathrm{S}}\right) < \mathrm{k}\right] \leq \binom{\mathrm{s}}{\mathrm{k}-1} (1-\delta)^{\mathrm{s}-\mathrm{k}+1}.$$

**Proof:** We sample a multi-set $S = \{i_1, \ldots, i_s\}$ uniformly at random. We say that an index $j \in [s]$ is *good* if either

$$\mathrm{rk}\left(\mathrm{G}|_{\{\mathrm{i}_1, \ldots, \mathrm{i}_{\mathrm{j}-1}\}}\right) = \mathrm{k} \quad \text{or} \quad \mathrm{rk}\left(\mathrm{G}|_{\{\mathrm{i}_1, \ldots, \mathrm{i}_{\mathrm{j}-1}\}}\right) < \mathrm{rk}\left(\mathrm{G}|_{\{\mathrm{i}_1, \ldots, \mathrm{i}_{\mathrm{j}}\}}\right).$$

Observe that $\mathrm{rk}\left(\mathrm{G}|_{\mathrm{S}}\right) < \mathrm{k}$ only if at most $k-1$ indices $j \in [s]$ are good. Combining the union bound over the choice of these indices with Lemma 1 completes the proof. ■

**Theorem 3** *Let $\delta > 8\delta' > 0$ be reals. There exists a constant $c_\delta$ such that for all integers $k \leq s \leq n$ where $s \geq c_\delta k$, for all prime powers $q$ and all $[n, k, \delta n]_q$ linear codes $C$, for a random choice of a multi-set $S$ we have*

$$\mathrm{Pr}_{\mathrm{S} \subseteq [\mathrm{n}] \mid |\mathrm{S}| = \mathrm{s}}\left[\mathrm{C}|_{\mathrm{S}} \textit{ is a } [s, k, \delta' s]_q \textit{ code}\right] \geq 1 - \exp_\delta(-\mathrm{k}).$$

**Proof:** We sample a multi-set $S = \{i_1, \ldots, i_s\}$ uniformly at random. We assume $s$ has the shape $ck\lceil 1/\delta \rceil$, where $ck$ is an integer. If not, then we omit some up to $\lceil 1/\delta \rceil - 1$ elements of $S$, and observe that such an omission has a negligible effect on the distance of the restriction. We arrange the set $S$ into $ck$ blocks of size $\lceil 1/\delta \rceil$,

$$S = \bigsqcup_{l=1}^{ck} B_l.$$

Set $\epsilon = 1/8$. Observe that $C|_S$ is not an $[s, k, \delta' s]_q$ code only if there exists a vector $\mathbf{x} \in \mathbb{F}_q^k$ such that $\mathbf{x}G|_S$ has weight below $\delta' s$. Thus there exist some $\delta' ck\lceil 1/\delta \rceil \leq 2\epsilon ck$ blocks that contain all non-zero coordinates of $\mathbf{x}G|_S$. Let $B$ be the union of these blocks. We necessarily have

$$\mathrm{rk}\left(\mathrm{G}|_{\mathrm{S}\setminus\mathrm{B}}\right) < \mathrm{k}.$$

Combining Lemma 2 with the union bound over the possible choices for $B$ we conclude that

$$\begin{gathered}\mathrm{Pr}_{\mathrm{S}\subseteq[\mathrm{n}]\ |\ |\mathrm{S}|=\mathrm{s}}\left[\mathrm{C}|_{\mathrm{S}} \text{ is not a } [s, k, \delta' s]_q \text{ code}\right] \leq \\ \binom{ck}{2\epsilon ck} \cdot \binom{(1-2\epsilon)ck\lceil 1/\delta \rceil}{k} \cdot (1-\delta)^{(1-2\epsilon)ck\lceil 1/\delta \rceil - k} \leq \\ \left(\frac{e}{2\epsilon}\right)^{2\epsilon ck} \cdot ((1-2\epsilon)c\lceil 1/\delta \rceil)^k \cdot \left(\frac{1}{1-\delta}\right)^k \cdot e^{-(1-2\epsilon)ck}.\end{gathered}$$

It is not hard to verify that for large enough $c$ the expression above decays exponentially fast with $k$. ■

Theorem 3 should be compared with an argument based on the Chernoff bound, that can be used to show that a random restriction to $O_\delta(k \log q)$ coordinates with high probability yields a code with linearly growing distance. That argument is much simpler and applies to both linear and non-linear codes. In contrast, Theorem 3 crucially uses linearity and gives a much better bound independent of $q$. It is possible to show that one cannot get a bound independent of $q$ for general non-linear codes.

### 3.1 Noisy interpolation of sparse polynomials

We now turn to interpolation of $k$-sparse polynomials. We assume that the collection of monomial degrees is known and that all of them are below $q-1$. We observe that in order to perform (non-noisy) interpolation it is sufficient to evaluate the unknown polynomial at (say) the first $k$ powers of the generator of the multiplicative group $\mathbb{F}_q^*$, and then solve the resulting Vandermonde system of linear equations. In what follows we show that performing *noisy* sparse interpolation requires only a constant (independent of $q$) multiplicative increase in the number of evaluation points. Our result is entailed by Theorem 3.

**Definition 4** *Let $q$ be a prime power, $\alpha > 0$ be a real, and $L$ be some collection of polynomials in $\mathbb{F}_q[x]$. We say that a set $S \subseteq F_q$ is an $\alpha$-noisy interpolating set for $L$, if any polynomial in $L$ can be uniquely recovered from its values on the set $S$, even after at most $\alpha|S|$ of these values have been corrupted arbitrarily.*

**Theorem 5** *Let $q$ be a prime power. Let $D \subseteq \{0, 1, \ldots, (1-\delta)q\}$ be an arbitrary set of size $k$, and $L \subseteq \mathbb{F}_q[x]$ be the set of polynomials whose monomial degrees belong to $D$; then for a random choice of a multi-set $S \subseteq \mathbb{F}_q$, $|S| = \Omega_\alpha(k)$ we have*

$$\mathrm{Pr}_{\mathrm{S}}\left[\mathrm{S} \textit{ is a } \delta/16\textit{-noisy interpol. set for } L\right] \geq 1 - \exp_\delta(-\mathrm{k}).$$

**Proof:** Let $C \subseteq \mathbb{F}_q^q$ be the linear space of evaluations of all polynomials in $L$. Clearly, $C$ is an $[q, k, \delta q]_q$ code. An application of Theorem 3 concludes the proof. ■

## 4 Locally decodable codes

Our results in this section follow by combining the polynomial view of matching vector codes developed in [9] with Theorem 3. We start with a brief introduction to matching vector codes.

**Definition 6** *A $q$-ary code $C : \mathbb{F}_q^k \to \mathbb{F}_q^N$ is said to be $(r, \delta, \epsilon)$-locally decodable if there exists a randomized decoding algorithm $\mathcal{A}$ such that*

1. *For all $\mathbf{x} \in \mathbb{F}_q^k$, $i \in [k]$ and $\mathbf{y} \in \mathbb{F}_q^N$ such that $d(C(\mathbf{x}), \mathbf{y}) \leq \delta N$ : $Pr[\mathcal{A}^{\mathbf{y}}(i) = \mathbf{x}(i)] \geq 1 - \epsilon$, where the probability is taken over the random coin tosses of the algorithm $\mathcal{A}$.*
2. *$\mathcal{A}$ makes at most $r$ queries to $\mathbf{y}$.*

**Definition 7** *Let $S \subseteq \mathbb{Z}_m \setminus \{0\}$. We say that families $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ and $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of vectors in $\mathbb{Z}_m^n$ form an $S$-matching family if the following two conditions are satisfied:*

- *For all $i \in [k]$, $(\mathbf{u}_i, \mathbf{v}_i) = 0$;*
- *For all $i, j \in [k]$ such that $i \neq j$, $(\mathbf{u}_j, \mathbf{v}_i) \in S$.*

We now show how one can obtain an MV code out of a matching family. We start with some notation.

- We assume that $q$ is a prime power, $m$ divides $q - 1$, and denote a subgroup of $\mathbb{F}_q^*$ of order $m$ by $\mathbb{C}_m$;
- We fix some generator $g$ of $\mathbb{C}_m$;
- For $\mathbf{w} \in \mathbb{Z}_m^n$, we define $g^{\mathbf{w}} \in \mathbb{C}_m^n$ by $\left(g^{\mathbf{w}(1)}, \ldots, g^{\mathbf{w}(n)}\right)$;
- For $\mathbf{w}, \mathbf{v} \in \mathbb{Z}_m^n$ we define the multiplicative line $M_{\mathbf{w},\mathbf{v}}$ through $\mathbf{w}$ in direction $\mathbf{v}$ to be the multi-set

$$M_{\mathbf{w},\mathbf{v}} = \left\{ g^{\mathbf{w}+\lambda\mathbf{v}} \mid \lambda \in \mathbb{Z}_m \right\}; \tag{2}$$

- For $\mathbf{u} \in \mathbb{Z}_m^n$, we define the monomial $\mathrm{mon}_{\mathbf{u}} \in \mathbb{F}_q[z_1, \ldots, z_n]$ by

$$\mathrm{mon}_{\mathbf{u}}(z_1, \ldots, z_n) = \prod_{\ell \in [n]} z_\ell^{\mathbf{u}(\ell)}. \tag{3}$$

### 4.1 Encoding/decoding framework for matching vector codes

Observe that for any $\mathbf{w}, \mathbf{u}, \mathbf{v} \in \mathbb{Z}_m^n$ and $\lambda \in \mathbb{Z}_m$ we have

$$\mathrm{mon}_{\mathbf{u}}\left(g^{\mathbf{w}+\lambda\mathbf{v}}\right) = g^{(\mathbf{u},\mathbf{w})}\left(g^{\lambda}\right)^{(\mathbf{u},\mathbf{v})}. \tag{4}$$

The formula above implies that the $M_{\mathbf{w},\mathbf{v}}$-evaluation of a monomial $\mathrm{mon}_{\mathbf{u}}$ is a $\mathbb{C}_m$-evaluation of a (univariate) monomial

$$g^{(\mathbf{u},\mathbf{w})} y^{(\mathbf{u},\mathbf{v})} \in \mathbb{F}_q[y]. \tag{5}$$

This observation is the foundation of all local decoders for matching vector codes. We now sketch encoding and decoding procedures. Let $\mathcal{U}, \mathcal{V}$ be an $S$-matching family in $\mathbb{Z}_m^n$.

**Encoding:** We encode a message $(\mathbf{x}(1), \ldots, \mathbf{x}(k)) \in \mathbb{F}_q^k$ by the $\mathbb{C}_m^n$-evaluation of the polynomial

$$F(z_1, \ldots, z_n) = \sum_{j=1}^{k} \mathbf{x}(j) \cdot \mathrm{mon}_{\mathbf{u}_j}(z_1, \ldots, z_n). \tag{6}$$

**Decoding:** The input to the decoder is a (corrupted) $\mathbb{C}_m^n$-evaluation of $F$ and an index $i \in [k]$.

1. The decoder picks $\mathbf{w} \in \mathbb{Z}_m^n$ uniformly at random;

2. The decoder recovers the noiseless restriction of $F$ to $M_{\mathbf{w},\mathbf{v}_i}$. To accomplish this the decoder queries the (corrupted) $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of $F$ at a certain number of locations.

To see that noiseless $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of $F$ uniquely determines $\mathbf{x}(i)$ note that by formulas (4), (5) and (6) the $M_{\mathbf{w},\mathbf{v}_i}$-evaluation of $F$ is a $\mathbb{C}_m$-evaluation of a polynomial

$$f(y) = \sum_{j=1}^{k} \mathbf{x}(j) \cdot g^{(\mathbf{u}_j,\mathbf{w})} y^{(\mathbf{u}_j,\mathbf{v}_i)} \in \mathbb{F}_q[y]. \tag{7}$$

Further observe that the properties of the $S$-matching family $\mathcal{U}, \mathcal{V}$ and (7) yield

$$f(y) = \mathbf{x}(i) \cdot g^{(\mathbf{u}_i,\mathbf{w})} + \sum_{s \in S} \left( \sum_{j\,:\,(\mathbf{u}_j,\mathbf{v}_i)=s} \mathbf{x}(j) \cdot g^{(\mathbf{u}_j,\mathbf{w})} \right) y^s. \tag{8}$$

It is evident from the above formula that the restriction of $F$ to a multiplicative line $M_{\mathbf{w},\mathbf{v}_i}$ yields a univariate polynomial $f(y)$ such that the set of monomial degrees of $f$ is in $S \cup \{0\}$ and

$$\mathbf{x}(i) = f(0)/g^{(\mathbf{u}_i,\mathbf{w})}. \tag{9}$$

### 4.2 Improved decoder for matching vector codes

We are now ready to state our main result for locally decodable codes.

**Theorem 8** *There exists $\delta > 0$, such that for all integers $t \geq 2$ and $k \geq 2^t$ there exists a prime power $q \leq \exp\left(t^{O(t)}\right)$, and a $q$-ary linear code encoding $k$-symbol messages to*

$$N = \exp\exp\left((\log k)^{1/t} (\log\log k)^{1-1/t} t \ln t\right)$$

*-symbol codewords that is $\left(O\left(2^t\right), \delta, 0.1\right)$-locally decodable.*

**Proof:** Our proof starts closely modeling the proof of [9, lemma 18]. We fix $t$ distinct primes close to $t \ln t$. We set $m = t^{O(t)}$ to be the product of these primes. We consider the Grolmusz family of matching vectors [13] mod $m$ in $\mathbb{Z}_m^n$ for an appropriate choice of the dimension $n$. We normalize the family to get an $S$-matching family where $|S| = 2^t$, and the largest element in $S$ is below $m/\ln t$. We then follow the strategy outlined in Section 4.1 to define the encoding procedure for the matching vector code.

At this point we depart from [9, lemma 18]. To recover the $i$-th message symbol, the local decoder in [9] relies only on the low degree of the restriction of the polynomial $F$ to a random multiplicative line $M_{\mathbf{w},\mathbf{v}_i}$. The decoder queries all $m = t^{O(t)}$ points on the line and performs noisy polynomial interpolation. Instead, we rely not only on low-degree but also on *sparsity* of the restriction of $F$ to $M_{\mathbf{w},\mathbf{v}_i}$. Specifically, we use the discussion in the end of Section 4.1 to conclude that possible restrictions of $F$ to $M_{\mathbf{w},\mathbf{v}_i}$ fall into a $\kappa = 2^t$-dimensional linear code in $\mathbb{F}_q^m$ that has good distance. By Theorem 3 the restriction of this code to a random subset of $O(\kappa)$ coordinates is also likely to have a good distance.

Thus to decode for $\mathbf{x}(i)$ we pick a random multiplicative line $M_{\mathbf{w},\mathbf{v}_i}$, query a randomly chosen set of its $O(2^t)$ points, recover the univariate polynomial $f(y)$ that is the most likely one given the observed values, and then obtain our candidate value of $\mathbf{x}(i)$ using formula (9). It is not hard to verify that by setting the value of $\delta$ appropriately small we can make the success probability of the decoder be arbitrary close to 1. ■

Note that in the theorem above it would be insufficient to use the naive result (about a restriction to $\kappa \log q$ coordinates) following from the Chernoff bound, since in the setting above $\log q = t^{O(t)} \gg 2^t = \kappa$.

## 5 Matrix rigidity

We start with a formal definition of a rigid matrix.

**Definition 9** *Let $q$ be a prime power and $m, n$, and $d$ be integers. Let $A$ be an $m \times n$ matrix over $\mathbb{F}_q$. We say that $A$ is $(r, d)$-rigid if for all matrices $A' \in \mathbb{F}_q^{m \times n}$ such that for all $i \in [m]$, $d(A_i, A'_i) \leq d$ we have* $\mathrm{rk(A')} > \mathrm{r}$.

It is not hard to verify that a matrix $A \in \mathbb{F}_q^{m \times n}$ is $(r, d)$-rigid if and only if for every linear space $L \subseteq \mathbb{F}_q^n$, $\dim L = r$ one of the rows of $A$ is more than $d$-far from $L$.

It is well-known that for any linear space $L$, $\dim L \leq (1-\epsilon)n$ there exists a point in $\mathbb{F}_q^n$ that is $\Omega(n)$-far from $L$. In fact, this statement can be shown via a simple greedy argument that uses only the cardinality of the set $L$. In the following two lemmas we show a similar statement for a point with $\{0,1\}$ coordinates. That argument however crucially relies on the fact that $L$ is a linear space.

**Lemma 10** *Let $q$ be a prime power and $L \subseteq \mathbb{F}_q^n$ be a linear subspace,* $\dim L = k$; *then*

$$\mathrm{Pr}_{\mathbf{x} \in \{0,1\}^{\mathrm{n}}} [\mathrm{x} \in \mathrm{L}] \leq \frac{1}{2^{\mathrm{n-k}}}. \tag{10}$$

**Proof:** Let $[n] = I \sqcup J$ be a disjoint partition of $[n]$, where $I$ is a set of information coordinates of $L$, i.e., $L|_I = \mathbb{F}_q^k$. Observe that a vector $\mathbf{x}_I \circ \mathbf{x}_J \in L$ is uniquely determined by $\mathbf{x}_I$. Thus for a random choice of a vector $\mathbf{x}_I \circ \mathbf{x}_J \in \{0,1\}^n$, there is at most $1/2^{n-k}$ chance that $\mathbf{x}_I \circ \mathbf{x}_J$ belongs to $L$. ■

**Lemma 11** *For every $\epsilon > 0$ there exists a $\delta > 0$, such that for all integers $n$, prime powers $q$, and linear spaces $L \subseteq \mathbb{F}_q^n$, $\dim L \leq (1-\epsilon)n$ there exists a point $\mathbf{x} \in \{0,1\}^n$ such that $d(L, \mathbf{x}) \geq \delta n$.*

**Proof:** Fix a linear space $L$. We now argue that a random point $\mathbf{x} \in \{0,1\}^n$ is far from $L$. Note that

$$\mathrm{Pr}_{\mathbf{x} \in \{0,1\}^{\mathrm{n}}} [\mathbf{x} \text{ is } \delta\text{-close to } L] = \mathrm{Pr}_{\mathbf{x} \in \{0,1\}^{\mathrm{n}}} [\exists \mathrm{S} \subseteq [\mathrm{n}], |\mathrm{S}| = (1-\delta)\mathrm{n} \text{ such that } \mathbf{x}|_{\mathrm{S}} \in \mathrm{L}|_{\mathrm{S}}].$$

We take a union bound over all $\binom{n}{\delta n}$ choices of the set $S$ and note that by Lemma 10 for any $S \subseteq [n]$, $|S| = (1-\delta)n$,

$$\mathrm{Pr}_{\mathbf{x} \in \{0,1\}^{\mathrm{n}}} [\mathbf{x}|_{\mathrm{S}} \in \mathrm{L}|_{\mathrm{S}}] \leq \frac{1}{2^{(1-\delta)\mathrm{n} - \dim \mathrm{L}|_{\mathrm{S}}}} \leq \frac{1}{2^{(\epsilon-\delta)\mathrm{n}}}.$$

Thus the probability that a randomly chosen $\mathbf{x} \in \{0,1\}^n$ is $\delta$-close to $L$ is at most $\binom{n}{\delta n} / 2^{(\epsilon-\delta)n}$ which is negligible when $\delta$ is sufficiently smaller than $\epsilon$. ■

We now proceed to our construction of explicit rigid matrices.

**Theorem 12** *Let $q$ be a prime power. For every $0 \leq d \leq O(n)$ there exists an $(n/2, d)$-rigid explicit matrix $A \in \mathbb{F}_q^{m \times n}$, with $m = 2^{O(d)} n/d$.*

**Proof:** Following our observation in the beginning of this section it suffices to construct an explicit set $A$ in $\mathbb{F}_q^n$ such that $|A| = m$, and for any linear space $L \subseteq \mathbb{F}_q^n$, $\dim L = n/2$ one of the points of $A$ is more than $d$-far from $L$.

By Lemma 11, there exists a constant $\delta > 0$ such that for any linear space $L \subseteq \mathbb{F}_q^n$ of dimension $n/2$ there is a point in $\{0,1\}^n$ that is more than $\delta n$-far from $L$.

To obtain the set $A$, split the coordinates into $n/d\lceil 1/\delta \rceil$ sets of size $d\lceil 1/\delta \rceil$ each, and in each set take all binary vectors with support on this set. $A$ consists of all these vectors. Note that every vector in $\{0,1\}^n$ is a sum of at most $\delta n/d$ vectors of our set $A$, whose size is $2^{O(d)}n/d$.

Now suppose that $L \subseteq \mathbb{F}_q^n$ is a linear space of dimension $n/2$ and every vector in $A$ is at most $d$-far from $L$. Then any vector of $A$ is a linear combination of a vector of $L$ and at most $d$ unit vectors. Hence any vector in $\{0,1\}^n$ is a linear combination of a vector of $L$ and at most $d(\delta n/d)$ unit vectors, contradicting the fact that in $\{0,1\}^n$ there exists a vector that is more than $\delta n$-far from $L$. ■

## 6 Conclusions

In this paper we introduced an iterative technique for analyzing the rank of a random minor of a matrix. We used this technique to establish a collection of results in coding theory and complexity theory. Below we comment on some problems that are left open by our work.

- Our results assert the existence of $O(k)$-sized noisy interpolating sets for $k$-sparse polynomials over finite fields. It would be very interesting to find such sets explicitly. Note that, as we have mentioned earlier, there are explicit $k$-sized sets for non-noisy sparse polynomial interpolation.
- Our results for locally decodable codes close the question regarding the size of the multiplicative increase in the query complexity of matching vector codes that is required to tolerate a constant fraction of errors. We show that (just as in the case with Reed Muller codes) a constant increase suffices. Further progress on matching vector codes requires better constructions of matching families of vectors modulo composites.
- In matrix rigidity it is a major challenge to improve upon the untouched minor argument at least for some range of parameters. One specific goal here could be to find an explicit $n \times n$ matrix $A$ whose rank stays above $0.99n$ under at most one alteration in every row. Note that a random matrix would tolerate up to $\Omega(n)$ alterations per row.

## References


[1] Noga Alon, Rina Panigrahy, and Sergey Yekhanin. Deterministic approximation algorithms for the nearest codeword problem. In *13th International Workshop on Randomization and Computation (RANDOM)*, volume 5687 of Lecture Notes in Computer Science, pages 339–351, 2009.

[2] Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. *Combinatorica*, 23:365–426, 2003.

[3] Laszlo Babai, Lance Fortnow, Leonid Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *23rd ACM Symposium on Theory of Computing (STOC)*, pages 21–31, 1991.

[4] Donald Beaver and Joan Feigenbaum. Hiding instances in multioracle queries. In *7th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 415 of Lecture Notes in Computer Science, pages 37–48. Springer, Berlin, Heidelberg, 1990.

[5] Avraham Ben-Aroya, Klim Efremenko, and Amnon Ta-Shma. Local list decoding with a constant number of queries. In *51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 715–722, 2010.

[6] Avraham Ben-Aroya, Klim Efremenko, and Amnon Ta-Shma. A note on amplifying the error-tolerance of locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR10-134, 2010.

[7] Michael Clausen, Andreas Dress, Johannes Grabmeier, and Marek Karpinski. On zero-testing and interpolation of $k$-sparse multivariate polynomials over finite fields. *Theoretical Computer Science*, 84:151–164, 1991.

[8] Zeev Dvir. On matrix rigidity and locally self-correctable codes. In *26th IEEE Computational Complexity Conference (CCC)*, pages 102–113, 2010.

[9] Zeev Dvir, Parikshit Gopalan, and Sergey Yekhanin. Matching vector codes. In *51st IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 705–714, 2010.

[10] Klim Efremenko. 3-query locally decodable codes of subexponential length. In *41st ACM Symposium on Theory of Computing (STOC)*, pages 39–44, 2009.

[11] Joel Friedman. A note on matrix rigidity. *Combinatorica*, 13:235–239, 1993.

[12] Dima Grigoriev, Marek Karpinski, and Michael Singer. Computational complexity of sparse rational interpolation. *SIAM Journal on Computing*, 23:1–11, 1994.

[13] Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica*, 20:71–86, 2000.

[14] Venkat Guruswami and Madhu Sudan. Improved decoding of Reed-Solomon and algebraic-geometric codes. *IEEE Transactions on Information Theory*, 45:1757–1767, 1999.

[15] Toshiya Itoh and Yasuhiro Suzuki. New constructions for query-efficient locally decodable codes of subexponential length. *IEICE Transactions on Information and Systems*, pages 263–270, 2010.

[16] Stasys Jukna. *Extremal combinatorics*. Springer, Berlin, Heidelberg, New York, 2001.

[17] Boris Kashin and Alexander Razborov. Improved lower bounds on the rigidity of Hadamard matrices. *Mathematical Notes*, 63:471–475, 1998.

[18] Richard Lipton. Efficient checking of computations. In *7th International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 415 of Lecture Notes in Computer Science, pages 207–215. Springer, Berlin, Heidelberg, 1990.

[19] Richard Lipton and Nisheeth Vishnoi. Deterministic identity testing for multivariate polynomials. In *14th Annual Symposium on Discrete Algorithms (SODA)*, pages 756–760, 2003.

[20] Satyanarayana Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity. *Journal of Computer and System Sciences*, 63:449–473, 2001.

[21] Satyanarayana Lokam. Complexity lower bounds using linear algebra. *Foundations and Trends in Theoretical Computer Science*, 4:1–155, 2009.

[22] Yishay Mansour. Randomized interpolation and approximation of sparse polynomials. *SIAM Journal on Computing*, 24:357–368, 1995.

[23] Yeow Meng Chee, Tao Feng, San Ling, Huaxiong Wang, and Liangfeng Zhang. Query-efficient locally decodable codes of subexponential length. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR10-173, 2010.

[24] Prasad Raghavendra. A note on Yekhanin's locally decodable codes. In *Electronic Colloquium on Computational Complexity (ECCC)*, TR07-016, 2007.

[25] Amin Shokrollahi, Daniel Speilman, and Voelker Stemann. A remark on matrix rigidity. *Information Processing Letters*, 64:283–285, 1997.

[26] Igor Shparlinski. Sparse polynomial approximation in finite fields. In *33rd ACM Symposium on Theory of Computing (STOC)*, pages 209–215, 2001.

[27] Igor Shparlinski and Arne Winterhof. Noisy interpolation of sparse polynomials in finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 16:307–317, 2005.

[28] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the XOR lemma. In *39th ACM Symposium on Theory of Computing (STOC)*, pages 537–546, 1999.

[29] Leslie Valiant. Graph-theoretic arguments in low level complexity. In *6th Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 162–176, 1977.

[30] Joachim von zur Gathen and Igor Shparilinski. Polynomial interpolation from multiples. In *15th Annual Symposium on Discrete Algorithms (SODA)*, pages 1125–1130, 2004.

[31] Kai Werther. The complexity of sparse polynomial interpolation over finite fields. *Applicable Algebra in Engineering, Communication and Computing*, 5:91–103, 1994.

[32] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM*, 55:1–16, 2008.

[33] Sergey Yekhanin. Locally decodable codes. *Foundations and trends in theoretical computer science*, 2011. to appear.