Restricted Isometry of Fourier Matrices and List Decodability of Random Linear Codes

Citation	Cheraghchi, Mahdi, Venkatesan Guruswami, and Ameya Velingker. "Restricted Isometry of Fourier Matrices and List Decodability of Random Linear Codes." SIAM Journal on Computing 42, no. 5 (October 2013): 1888–1914. © 2013, Society for Industrial and Applied Mathematics
As Published	http://dx.doi.org/10.1137/120896773
Publisher	Society for Industrial and Applied Mathematics
Version	Final published version
Citable link	http://hdl.handle.net/1721.1/86093
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

SIAM J. Comput.
Vol. 42, No. 5, pp. 1888–1914

RESTRICTED ISOMETRY OF FOURIER MATRICES AND LIST DECODABILITY OF RANDOM LINEAR CODES[*]

MAHDI CHERAGHCHI[†], VENKATESAN GURUSWAMI[‡], AND AMEYA VELINGKER[§]

Abstract. We prove that a random linear code over $\mathbb{F}_q$, with probability arbitrarily close to 1, is list decodable at radius $1 - 1/q - \epsilon$ with list size $L = O(1/\epsilon^2)$ and rate $R = \Omega_q(\epsilon^2/(\log^3(1/\epsilon)))$. Up to the polylogarithmic factor in $1/\epsilon$ and constant factors depending on q, this matches the lower bound $L = \Omega_q(1/\epsilon^2)$ for the list size and upper bound $R = O_q(\epsilon^2)$ for the rate. Previously only existence (and not abundance) of such codes was known for the special case $q = 2$ (Guruswami et al., 2002). In order to obtain our result, we employ a relaxed version of the well-known Johnson bound on list decoding that translates the *average* Hamming distance between codewords to list decoding guarantees. We furthermore prove that the desired average-distance guarantees hold for a code provided that a natural complex matrix encoding the codewords satisfies the restricted isometry property with respect to the Euclidean norm. For the case of random binary linear codes, this matrix coincides with a random submatrix of the Hadamard–Walsh transform matrix that is well studied in the compressed sensing literature. Finally, we improve the analysis of Rudelson and Vershynin (2008) on the number of random frequency samples required for exact reconstruction of k-sparse signals of length N. Specifically, we improve the number of samples from $O(k \log(N) \log^2(k)(\log k + \log\log N))$ to $O(k \log(N) \cdot \log^3(k))$. The proof involves bounding the expected supremum of a related Gaussian process by using an improved analysis of the metric defined by the process. This improvement is crucial for our application in list decoding.

Key words. combinatorial list decoding, compressed sensing, Gaussian processes

AMS subject classifications. 05B40, 46B09, 60G15, 94A20, 94B65

DOI. 10.1137/120896773

1. Introduction.

This work is motivated by the list decodability properties of random linear codes for correcting a large fraction of errors, approaching the information-theoretic maximum limit. We prove a near-optimal bound on the rate of such codes by making a connection to and establishing improved bounds on the restricted isometry property (RIP) of random submatrices of Hadamard matrices.

A q-ary error correcting code $\mathcal{C}$ of block length n is a subset of $[q]^n$, where $[q]$ denotes any alphabet of size q. The rate of such a code is defined to be $(\log_q |\mathcal{C}|)/n$. A good code $\mathcal{C}$ should be large (rate bounded away from 0) and have its elements (codewords) well spread out. The latter property is motivated by the task of recovering a codeword $c \in \mathcal{C}$ from a noisy version r of it that differs from c in a bounded number of coordinates. Since a random string $r \in [q]^n$ will differ from c on an expected $(1-1/q)n$ positions, the information-theoretically maximum fraction of errors one can correct is bounded by the limit $(1-1/q)$. In fact, when the fraction of errors exceeds $\frac{1}{2}(1-1/q)$,

[*] Received by the editors October 29, 2012; accepted for publication (in revised form) July 15, 2013; published electronically October 1, 2013. An extended abstract on this work is to appear in *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA 2013).

http://www.siam.org/journals/sicomp/42-5/89677.html

[†] Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA 02139 (mahdi@csail.mit.edu). This research was done while the author was with the Computer Science Department of Carnegie Mellon University. This author's work was supported in part by Swiss National Science Foundation research grant PBELP2-133367.

[‡] Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213 (guruswami@cmu.edu). This author's research was supported in part by a Packard Fellowship and NSF CCF-0963975.

[§] Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213 (avelingk@cs.cmu.edu). This author's research was supported in part by NSF CCF-0963975.

it is not possible to unambiguously identify the close-by codeword to the noisy string r (unless the code has very few codewords, i.e., a rate approaching zero).

In the model of list decoding, however, recovery from a fraction of errors approaching the limit $(1 - 1/q)$ becomes possible. Under list decoding, the goal is to recover a small list of all codewords of $\mathcal{C}$ differing from an input string r in at most ρn positions, where ρ is the error fraction (our interest in this paper being the case when ρ is close to $1 - 1/q$). This requires that $\mathcal{C}$ have the following sparsity property, called (ρ, L)-*list decodability*, for some small L : for every $r \in [q]^n$, there are at most L codewords within Hamming distance ρn from r. We will refer to the parameter L as the "list size"—it refers to the maximum number of codewords that the decoder may output when correcting a fraction ρ of errors. Note that (ρ, L)-list decodability is a strictly combinatorial notion and does not promise an efficient algorithm to compute the list of close-by codewords. In this paper, we only focus on this combinatorial aspect and study a basic trade-off between between ρ, L, and the rate for the important class of random linear codes, when $\rho \to 1 - 1/q$. We describe the prior results in this direction and state our results next.

For integers $q, L \geq 2$, a random q-ary code of rate $R = 1 - h_q(\rho) - 1/L$ is (ρ, L)-list decodable with high probability. Here $h_q\colon [0, 1 - 1/q] \to [0, 1]$ is the q-ary entropy function: $h_q(x) = x \log_q(q - 1) - x \log_q x - (1 - x) \log_q(1 - x)$. This follows by a straightforward application of the probabilistic method, based on a union bound over all centers $r \in [q]^n$ and all $(L + 1)$-element subsets S of codewords that all codewords in S lie in the Hamming ball of radius ρn centered at r. For $\rho = 1 - 1/q - \epsilon$, where we think of q as fixed and $\epsilon \to 0$, this implies that a random code of rate $\Omega_q(\epsilon^2)$ is $(1 - 1/q - \epsilon, O_q(1/\epsilon^2))$-list decodable. (Here and below, the notation Ω_q and O_q hides constant factors that depend only on q.)

Understanding list decodable codes at the extremal radii $\rho = 1 - 1/q - \epsilon$, for small ϵ, is of particular significance mainly due to numerous applications that depend on this regime of parameters. For example, one can mention hardness amplification of Boolean functions [29], construction of hardcore predicates from one-way functions [14], construction of pseudorandom generators [29] and randomness extractors [30], inapproximability of NP witnesses [24], and approximating the VC dimension [26]. Moreover, *linear* list-decodable codes are further appealing due to their symmetries, succinct description, and efficient encoding. For some applications, linearity of list decodable codes is of crucial importance. For example, the black-box reduction from list decodable codes to capacity achieving codes for additive noise channels in [20] or certain applications of Trevisan's extractor [30] (e.g., [10, sections 3.6, 5.2]) rely on linearity of the underlying list decodable code. Furthermore, list decoding of linear codes features an interplay between linear subspaces and Hamming balls and their intersection properties, which is of significant interest from a combinatorial perspective.

This work is focused on random *linear* codes, which are subspaces of $\mathbb{F}_q^n$, where $\mathbb{F}_q$ is the finite field with q elements. A random linear code $\mathcal{C}$ of rate R is sampled by picking $k = Rn$ random vectors in $\mathbb{F}_q^n$ and letting $\mathcal{C}$ be their $\mathbb{F}_q$-span. Since the codewords of $\mathcal{C}$ are now not all independent (in fact they are not even 3-wise independent), the above naive argument only proves the (ρ, L)-list decodability property for codes of rate $1 - h_q(\rho) - 1/\log_q(L + 1)$ [33].[1] For the setting $\rho = 1 - 1/q - \epsilon$,

[1] The crux of the argument is that any L nonzero vectors in $\mathbb{F}_q^k$ must have a subset of $\log_q(L+1)$ linearly independent vectors, and these are mapped independently by a random linear code. This allows one to effectively substitute $\log_q(L + 1)$ in the place of L in the argument for fully random codes.

this implies a list size bound of $\exp(O_q(1/\epsilon^2))$ for random linear codes of rate $\Omega_q(\epsilon^2)$, which is exponentially worse than for random codes. Understanding if this exponential discrepancy between general and linear codes is inherent was raised as an open question by Elias [13]. Despite much research, the exponential bound was the best known for random linear codes (except for the case of $q = 2$, and even for $q = 2$ only an existence result was known; see the related results section below for more details).

Our main result in this work closes this gap between random linear and random codes up to polylogarithmic factors in the rate. We state a simplified version of the main theorem (Theorem 3.8) below.

THEOREM 1.1 (main, simplified). *Let q be a prime power, and let $\epsilon > 0$ be a constant parameter. Then for some constant $a_q > 0$ only depending on q and all large enough integers n, a random linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate $a_q\epsilon^2/\log^3(1/\epsilon)$ is $(1 - 1/q - \epsilon, O(1/\epsilon^2))$-list decodable with probability at least 0.99. (One can take $a_q = \Omega(1/\log^4 q)$.)*

We remark that both the rate and the list size are close to optimal for list decoding from a $(1 - 1/q - \epsilon)$ fraction of errors. For rate, this follows from the fact the q-ary "list decoding capacity" is given by $1 - h_q(\rho)$, which is $O_q(\epsilon^2)$ for $\rho = 1 - 1/q - \epsilon$. For list size, a lower bound of $\Omega_q(1/\epsilon^2)$ is known—this follows from [3] for $q = 2$ and was shown for all q in [22, 4]. We have also assumed that the alphabet size q is fixed and have not attempted to obtain the best possible dependence of the constants on the alphabet size.

1.1. Related results. We now discuss some other previously known results concerning list decodability of random linear codes.

First, it is well known that a random linear code of rate $\Omega_q(\epsilon^4)$ is $(1 - 1/q - \epsilon, O(1/\epsilon^2))$-list decodable with high probability. This follows by combining the Johnson bound for list decoding (see, for example, [21]) with the fact that such codes lie on the Gilbert–Varshamov bound and have relative distance $1-1/q-\epsilon^2$ with high probability. This result gets the correct quadratic dependence in list size, but the rate is worse.

Second, for the case of $q = 2$, the existence of (ρ, L)-list decodable binary linear codes of rate $1 - h(\rho) - 1/L$ was proved in [18]. For $\rho = 1/2 - \epsilon$, this implies the existence of binary linear codes of rate $\Omega(\epsilon^2)$ list decodable with list size $O(1/\epsilon^2)$ from an error fraction $1/2 - \epsilon$. This matches the bounds for random codes and is optimal up to constant factors. However, there are two shortcomings with this result: (i) it works only for $q = 2$ (the proof makes use of this in a crucial way, and extensions of the proof to larger q have been elusive), and (ii) the proof is based on the semi-random method. It only shows the existence of such a code while failing to give any sizeable lower bound on the probability that a random linear code has the claimed list decodability property.

Motivated by this state of affairs, in [17], the authors proved that a random q-ary linear code of rate $1 - h_q(\rho) - C_{\rho,q}/L$ is (ρ, L)-list decodable with high probability for some $C_{\rho,q} < \infty$ that depends on ρ, q. This matches the result for completely random codes up to the leading constant $C_{\rho,q}$ in front of $1/L$. Unfortunately, for $\rho = 1 - 1/q - \epsilon$, the constant $C_{\rho,q}$ depends exponentially[2] on $1/\epsilon$. Thus, this result

[2] The constant $C_{\rho,q}$ depends exponentially on $1/\delta_\rho$, where $q^{-\delta_\rho n}$ is an upper bound on the probability that two random vectors in $\mathbb{F}_q^n$ of relative Hamming weight at most ρ, chosen independently and uniformly among all possibilities, sum up (over $\mathbb{F}_q^n$) to a vector of Hamming weight at most ρ. When $\rho = 1 - 1/q - \epsilon$, we have $\delta_\rho = \Theta_q(\epsilon^2)$, which makes the list size exponentially large.

only implies an exponential list size in $1/\epsilon$, as opposed to the optimal $O(1/\epsilon^2)$ that we seek.

Summarizing, for random linear codes to achieve a polynomial in $1/\epsilon$ list size bound for error fraction $1 - 1/q - \epsilon$, the best lower bound on rate was $\Omega(\epsilon^4)$. We are able to show that random linear codes achieve a list size growing quadratically in $1/\epsilon$ for a rate of $\tilde{\Omega}(\epsilon^2)$. One downside of our result is that we do not get a probability bound of $1 - o(1)$ but get only $1 - \gamma$ for any desired constant $\gamma > 0$. (Essentially our rate bound degrades by a $\log(1/\gamma)$ factor.)

Finally, there are also some results showing limitations on list decodability of random codes. It is known that both random codes and random linear codes of rate $1 - h_q(\rho) - \eta$ are, with high probability, *not* $(\rho, c_{\rho,q}/\eta)$-list decodable [28, 19]. For arbitrary (not necessarily random) codes, the best lower bound on list size is $\Omega(\log(1/\eta))$ [3, 19].

Remark 1.2. We note that subsequent to the initial draft of this paper, an improved version of our coding result was obtained in [32], where it is shown that the rate of a random linear code can be improved to $\Omega(\epsilon^2/\log(q))$ while achieving $((1-1/q)(1-\epsilon), O(1/\epsilon^2))$-list decodability with probability $1-o(1)$, thereby obtaining the optimal dependence of rate on ϵ. While [32] does make use of the simplex encoding technique used here, it bypasses the use of RIP-2 and instead controls a related L_1 norm to achieve a simpler proof of the list decodability result. However, as a result, it does not improve the number of row samples of a DFT matrix needed to obtain RIP-2, a question that is interesting in its own right.

1.2. Proof technique. The proof of our result uses a different approach from the earlier works on list decodability of random linear codes [33, 13, 18, 17]. Our approach consists of three steps.

Step 1. Our starting point is a relaxed version of the Johnson bound for list decoding that requires only the *average* pairwise distance of L codewords to be large (where L is the target list size), instead of the minimum distance of the code.

Technically, this extension is easy and pretty much follows by inspecting the proof of the Johnson bound. This has recently been observed for the binary case by Cheraghchi [11]. Here, we give a proof of the relaxed Johnson bound for a more general setting of parameters and apply it in a setting where the usual Johnson bound is insufficient. Furthermore, as a side application, we show how the average version can be used to bound the list decoding radius of codes which do not have too many codewords close to any codeword—such a bound was shown via a different proof in [16], where it was used to establish the list decodability of binary Reed–Muller codes up to their distance.

Step 2. Prove that the L-wise average distance property of random linear codes is implied by the order L restricted isometry property (RIP-2) of random submatrices of the Hadamard matrix (or in general, matrices related to the DFT).

This part is also easy technically, and our contribution lies in making this connection between restricted isometry and list decoding. The RIP has received much attention lately due to its relevance to compressed sensing (cf. [5, 6, 7, 8, 12]) and is also connected to the Johnson–Lindenstrauss dimension reduction lemma [2, 1, 23]. Our work shows another interesting application of this concept.

Step 3. Prove the needed RIP of the matrix obtained by sampling rows of the Hadamard matrix.

This is the most technical part of our proof. Let us focus on $q = 2$ for simplicity, and let H be the $N \times N$ Hadamard (DFT) matrix with $N = 2^n$, whose (x, y)th entry is

$(-1)^{\langle x,y\rangle}$ for $x, y \in \{0,1\}^n$. We prove that (the scaled version of) a random submatrix of H formed by sampling a subset of $m = O(k \log^3 k \log N)$ rows of H satisfies RIP of order k with probability 0.99. This means that every k columns of this sampled matrix M are nearly orthogonal—formally, every $m \times k$ submatrix of M has all its k singular values close to 1.

For random matrices $m\times N$ with independent and identically distributed Gaussian or (normalized) ± 1 entries, it is relatively easy to prove RIP-2 of order k when $m = O(k \log N)$ [2]. Proving such a bound for submatrices of the DFT matrix (as conjectured in [27]) has been an open problem for many years. The difficulty is that the entries within a row are no longer independent and not even triplewise independent. The best proven upper bound on m for this case was $O(k \log^2 k(\log k + \log\log N) \log N)$, improving an earlier upper bound $O(k \log^6 N)$ of Candès and Tao [8]. We improve the bound to $O(k \log^3 k \log N)$—the key gain is that we do *not* have the $\log\log N$ factor. This is crucial for our list decoding connection, as the rate of the code associated with the matrix will be $(\log N)/m$, which would be $o(1)$ if $m = \Omega(\log N \log\log N)$. We will take $k = L = \Theta(1/\epsilon^2)$ (the target list size), and the rate of the random linear code will be $\Omega(1/(k \log^3 k))$, giving the bounds claimed in Theorem 1.1. We remark that any improvement of the RIP bound toward the information-theoretic limit $m = \Omega(k \log(N/k))$, a challenging open problem, would immediately translate into an improvement on the list decoding rate of random linear codes via our reductions.

Our RIP-2 proof for row-subsampled DFT matrices proceeds along the lines of [27] and is based on upper bounding the expectation of the supremum of a certain *Gaussian process* [25, Chapter 11]. The index set of the Gaussian process is $\mathcal{B}_2^{k,N}$, the set of all k-sparse unit vectors in $\mathbb{R}^N$, and the Gaussian random variable G_x associated with $x \in \mathcal{B}_2^{k,N}$ is a Gaussian linear combination of the squared projections of x on the rows sampled from the DFT matrix (in the binary case these are just squared Fourier coefficients).[3] The key to analyzing the Gaussian process is an understanding of the associated (pseudo-)metric X on the index set, defined by $\|x-x'\|_X^2 = \mathbb{E}_G|G_x - G_{x'}|^2$. This metric is difficult to work with directly, so we upper bound distances under X in terms of distances under a different metric X'. The principal difference in our analysis compared to [27] is in the choice of X'—instead of the max norm used in [27], we use an L_p norm for large finite p applied to the sampled Fourier coefficients. We then estimate the covering numbers for X' and use Dudley's theorem to bound the supremum of the Gaussian process.

It is worth pointing out that, as we prove in this work, for low-rate random linear codes the average-distance quantity discussed in Step 1 above is substantially larger than the minimum distance of the code. This allows the relaxed version of the Johnson bound to attain better bounds than what the standard (minimum-distance-based) Johnson bound would obtain on list decodability of random linear codes. While explicit examples of linear codes surpassing the standard Johnson bound are already known in the literature (see [15] and the references therein), a by-product of our result is that in fact *most* linear codes (at least in the low-rate regime) surpass the standard Johnson bound. However, an interesting question is to see whether there are codes

[3] We should remark that our setup of the Gaussian process is slightly different from [27], where the index set is k-element subsets of $[N]$ and the associated Gaussian random variable is the spectral norm of a random matrix. Moreover, in [27] the number of rows of the subsampled DFT matrix is a random variable concentrating around its expectation, contrary to our case, where it is a fixed number. We believe that the former difference in our setup may make the proof accessible to a broader audience.

that are list decodable even beyond the relaxed version of the Johnson bound studied in this work.

Organization of the paper. The rest of the paper is organized as follows. After fixing some notation, in section 2 we prove the average-case Johnson bound that relates a lower bound on average pairwise distances of subsets of codewords in a code to list decoding guarantees on the code. We also show, in section 2.3, an application of this bound on proving list decodability of "locally sparse" codes, which is of independent interest and simplifies some earlier list decoding results. In section 3, we prove our main theorem on list decodability of random linear codes by demonstrating a reduction from RIP-2 guarantees of DFT-based complex matrices to average distance of random linear codes, combined with the Johnson bound. Finally, the RIP-2 bounds on matrices related to random linear codes are proved in section 4.

Notation. Throughout the paper, we will be interested in list decodability of q-ary codes. We will denote an alphabet of size q by $[q] := \{1, \ldots, q\}$. For linear codes, the alphabet will be $\mathbb{F}_q$, the finite field with q elements (when q is a prime power). However, whenever there is a need to identify $\mathbb{F}_q$ with $[q]$ and vice versa (for example, to form the simplex encoding in Definition 2.4), we implicitly assume a fixed, but arbitrary, bijection between the two sets.

We use the notation $\mathbf{i} := \sqrt{-1}$. When $f \leq Cg$ (resp., $f \geq Cg$) for some absolute constant $C > 0$, we use the shorthand $f \lesssim g$ (resp., $f \gtrsim g$). We use the notation $\log(\cdot)$ when the base of logarithm is not of significance (e.g., $f \lesssim \log x$). Otherwise the base is subscripted as in $\log_b(x)$. The natural logarithm is denoted by $\ln(\cdot)$.

For a matrix M and a multiset of rows T, define M_T to be the matrix with $|T|$ rows, formed by the rows of M picked by T (in some arbitrary order). Each row in M_T may be repeated for the appropriate number of times specified by T.

2. Average-distance-based Johnson bound.

In this section, we show how the average pairwise distances between subsets of codewords in a q-ary code translate into list decodability guarantees on the code.

Recall that the relative Hamming distance between strings $x, y \in [q]^n$, denoted $\delta(x, y)$, is defined to be the fraction of positions i for which $x_i \neq y_i$. The relative distance of a code $\mathcal{C}$ is the minimum value of $\delta(x, y)$ over all pairs of codewords $x \neq y \in \mathcal{C}$. We define list decodability as follows.

DEFINITION 2.1. *A code $\mathcal{C} \subseteq [q]^n$ is said to be (ρ, ℓ)-list decodable if for all $y \in [q]^n$, the number of codewords of $\mathcal{C}$ within relative Hamming distance less than ρ is at most ℓ.*[4]

The following definition captures a crucial function that allows one to generically pass from distance property to list decodability.

DEFINITION 2.2 (Johnson radius). *For an integer $q \geq 2$, the Johnson radius function $J_q : [0, 1 - 1/q] \to [0, 1]$ is defined by*

$$J_q(x) := \frac{q-1}{q}\left(1 - \sqrt{1 - \frac{qx}{q-1}}\right).$$

The well-known Johnson bound in coding theory states that a q-ary code of relative distance δ is $(J_q(\delta - \delta/L), L)$-list decodable (see, for instance, [21]). Below we prove a version of this bound which does not need every pair of codewords to be far apart but instead works when the average distance of every set of codewords is large. The proof of this version of the Johnson bound is a simple modification of earlier

[4] We require that the radius is strictly less than ρ instead of at most ρ for convenience.

proofs, but working with this version is a crucial step in our near-tight analysis of the list decodability of random linear codes.

THEOREM 2.3 (average-distance Johnson bound). *Let $\mathcal{C} \subseteq [q]^n$ be a q-ary code and $L \geq 2$ an integer. If the average pairwise relative Hamming distance of every subset of L codewords of $\mathcal{C}$ is at least δ, then $\mathcal{C}$ is $(J_q(\delta - \delta/L), L-1)$-list decodable.*

Thus, if one is interested in a bound for list decoding with list size L, it is enough to consider the average pairwise Hamming distance of subsets of L codewords.

2.1. Geometric encoding of q-ary symbols.

We will give a geometric proof of the above result. For this purpose, we will map vectors in $[q]^n$ to complex vectors and argue about the inner products of the resulting vectors.

DEFINITION 2.4 (simplex encoding). *The simplex encoding maps $x \in [q]$ to a vector $\varphi(x) \in \mathbb{C}^{q-1}$. The coordinate positions of this vector are indexed by the elements of $[q-1] := \{1, 2, \ldots, q-1\}$. Namely, for every $\alpha \in [q-1]$, we define $\varphi(x)(\alpha) := \omega^{x\alpha}$, where $\omega = e^{2\pi \mathbf{i}/q}$ is the primitive qth complex root of unity.*

For complex vectors $v = (v_1, v_2, \ldots, v_m)$ and $w = (w_1, w_2, \ldots, w_m)$, we define their inner product $\langle v, w\rangle = \sum_{i=1}^m v_i w_i^*$. From the definition of the simplex encoding, the following immediately follows:

$$\langle \varphi(x), \varphi(y)\rangle = \begin{cases} q-1 & \text{if } x = y, \\ -1 & \text{if } x \neq y. \end{cases} \tag{2.1}$$

We can extend the above encoding to map elements of $[q]^n$ into $\mathbb{C}^{n(q-1)}$ in the natural way by applying this encoding to each coordinate separately. From the above inner product formula, it follows that for $x, y \in [q]^n$ we have

$$\langle \varphi(x), \varphi(y)\rangle = (q-1)n - q\delta(x,y)n. \tag{2.2}$$

Similarly, we overload the notation to matrices with entries over $[q]$. Let M be a matrix in $[q]^{n\times N}$. Then, $\varphi(M)$ is an $n(q-1) \times N$ complex matrix obtained from M by replacing each entry with its simplex encoding, considered as a column complex vector.

Finally, we extend the encoding to *sets* of vectors (i.e., codes) as well. For a set $\mathcal{C} \subseteq [q]^n$, $\varphi(\mathcal{C})$ is defined as a $(q-1)n \times |\mathcal{C}|$ matrix with columns indexed by the elements of $\mathcal{C}$, where the column corresponding to each $c \in \mathcal{C}$ is set to be $\varphi(c)$.

2.2. Proof of average-distance Johnson bound.

We now prove the Johnson bound based on average distance.

Proof of Theorem 2.3. Suppose $\{c_1, c_2, \ldots, c_L\} \subseteq [q]^n$ are such that their average pairwise relative distance is at least δ, i.e.,

$$\sum_{1 \leq i < j \leq L} \delta(c_i, c_j) \geq \delta \cdot \binom{L}{2}. \tag{2.3}$$

We will prove that $c_1, c_2, \ldots, c_L$ cannot all lie in a Hamming ball of radius less than $J_q(\delta - \delta/L)n$. Since every subset of L codewords of $\mathcal{C}$ satisfies (2.3), this will prove that $\mathcal{C}$ is $(J_q(\delta - \delta/L), L-1)$-list decodable.

Suppose, for contradiction, that there exists $c_0 \in [q]^n$ such that $\delta(c_0, c_i) \leq \rho$ for $i = 1, 2, \ldots, L$ and some $\rho < J_q(\delta - \delta/L)$. Recalling the definition of $J_q(\cdot)$, note that the assumption about ρ implies

$$\left(1 - \frac{q\rho}{q-1}\right)^2 > 1 - \frac{q\delta}{q-1} + \frac{q}{q-1}\frac{\delta}{L}\ . \tag{2.4}$$

For $i = 1, 2, \ldots, L$, define the vector $v_i = \varphi(c_i) - \beta\varphi(c_0) \in \mathbb{C}^{n(q-1)}$ for some parameter β to be chosen later. By (2.2) and the assumptions about $c_0, c_1, \ldots, c_L$, we have $\langle \varphi(c_i), \varphi(c_0)\rangle \geq (q-1)n - q\rho n$, and $\sum_{1\leq i<j\leq L} \langle \varphi(c_i), \varphi(c_j)\rangle \leq \binom{L}{2}\big((q-1)n - q\delta n\big)$. We have

$$\begin{aligned}0 &\leq \left\langle \sum_{i=1}^{L} v_i,\ \sum_{i=1}^{L} v_i \right\rangle = \sum_{i=1}^{L} \langle v_i, v_i\rangle + 2 \cdot \sum_{1\leq i<j\leq L} \langle v_i, v_j\rangle \\ &\leq L\big(n(q-1) + \beta^2 n(q-1) - 2\beta(n(q-1) - q\rho n)\big) \\ &\quad + L(L-1)\big(n(q-1) - q\delta n + \beta^2 n(q-1) - 2\beta(n(q-1) - q\rho n)\big) \\ &= L^2 n(q-1)\left(\frac{q}{q-1}\frac{\delta}{L} + \left(1 - \frac{q\delta}{q-1} + \beta^2 - 2\beta\Big(1 - \frac{q\rho}{q-1}\Big)\right)\right).\end{aligned}$$

Picking $\beta = 1 - \frac{q\rho}{q-1}$ and recalling (2.4), we see that the above expression is negative, a contradiction. □

2.3. An application: List decodability of Reed–Muller and locally sparse codes.

Our average-distance Johnson bound implies the following combinatorial result for the list decodability of codes that have few codewords in a certain vicinity of every codeword. The result allows one to translate a bound on the number of codewords in balls centered at codewords to a bound on the number of codewords in an arbitrary Hamming ball of smaller radius. An alternate proof of the below bound (using a "deletion" technique) was given by Gopalan, Klivans, and Zuckerman [16], who used it to argue the list decodability of (binary) Reed–Muller codes up to their relative distance. A mild strengthening of the deletion lemma was later used in [15] to prove combinatorial bounds on the list decodability of tensor products and interleavings of binary linear codes.

LEMMA 2.5. *Let $q \geq 2$ be an integer and $\eta \in (0, 1 - 1/q]$. Suppose $\mathcal{C}$ is a q-ary code such that for every $c \in \mathcal{C}$, there are at most A codewords of relative distance less than η from c (including c itself). Then, for every positive integer $L \geq 2$, $\mathcal{C}$ is $(J_q(\eta - \eta/L), AL - 1)$-list decodable.*

Note that setting $A = 1$ above gives the usual Johnson bound for a code of relative distance at least η.

Proof. We will lower bound the average pairwise relative distance of every subset of AL codewords of $\mathcal{C}$ and then apply Theorem 2.3.

Let $c_1, c_2, \ldots, c_{AL}$ be distinct codewords of $\mathcal{C}$. For $i = 1, 2, \ldots, AL$, the sum of relative distances of c_j, $j \neq i$, from c_i is at least $(AL - A)\eta$ since there are at most A codewords at relative distance less than η from c_i. Therefore

$$\frac{1}{\binom{AL}{2}} \cdot \sum_{1\leq i<j\leq AL} \delta(c_i, c_j) \geq \frac{AL \cdot (AL - A)\eta}{AL(AL-1)} = \frac{A(L-1)}{AL-1}\eta\ .$$

Setting $\eta' = \frac{A(L-1)\eta}{AL-1}$, Theorem 2.3 implies that $\mathcal{C}$ is $(J_q(\eta' - \frac{\eta'}{AL}), AL - 1)$-list decodable. But $\eta' - \frac{\eta'}{AL} = \eta - \eta/L$, so the claim follows. □

3. Proof of the list decoding result.

In this section, we prove our main result on list decodability of random linear codes. The main idea is to use the RIP of complex matrices arising from random linear codes for bounding average pairwise distances of subsets of codewords. Combined with the average-distance-based Johnson bound shown in the previous section, this proves the desired list decoding bounds. The RIP-2 condition that we use in this work is defined as follows.

Definition 3.1. *We say that a complex matrix $M \in \mathbb{C}^{m \times N}$ satisfies RIP-2 of order k with constant δ if, for any k-sparse vector $x \in \mathbb{C}^N$, we have*[5]

$$(1-\delta)\|x\|_2^2 \leq \|Mx\|_2^2 \leq (1+\delta)\|x\|_2^2.$$

Generally we think of δ as a small positive constant, say, $\delta = 1/2$.

Since we will be working with list decoding radii close to $1 - 1/q$, we derive a simplified expression for the Johnson bound in this regime, namely, the following.

Theorem 3.2. *Let $\mathcal{C} \subseteq [q]^n$ be a q-ary code and $L \geq 2$ an integer. If the average pairwise relative Hamming distance of every subset of L codewords of $\mathcal{C}$ is at least $(1-1/q)(1-\epsilon)$, then $\mathcal{C}$ is $((1-1/q)(1-\sqrt{\epsilon + 1/L}), L-1)$-list decodable.*

Proof. The proof is nothing but a simple manipulation of the bound given by Theorem 2.3. Let $\delta := (1-1/q)(1-\epsilon)$. Theorem 2.3 implies that $\mathcal{C}$ is $(J_q(\delta(1 - 1/L)), L-1)$-list decodable. Now,

$$\begin{aligned} J_q(\delta(1-1/L)) &= \frac{q-1}{q}\left(1 - \sqrt{1 - \frac{q}{q-1} \cdot \frac{q-1}{q}(1-\epsilon)\Big(1-\frac{1}{L}\Big)}\right) \\ &= \frac{q-1}{q}\left(1-\sqrt{\epsilon + \frac{1}{L} - \frac{\epsilon}{L}}\right) \geq \frac{q-1}{q}\left(1 - \sqrt{\epsilon + \frac{1}{L}}\right). \qquad \square \end{aligned}$$

In order to prove lower bounds on average distance of random linear codes, we will use the simplex encoding of vectors (Definition 2.4), along with the following simple geometric lemma.

Lemma 3.3. *Let $c_1, \ldots, c_L \in [q]^n$ be q-ary vectors. Then, the average pairwise distance δ between these vectors satisfies*

$$\delta := \sum_{1 \leq i < j \leq L} \delta(c_i, c_j) \Big/ \binom{L}{2} = \frac{L^2(q-1)n - \left\| \sum_{i \in [L]} \varphi(c_i) \right\|_2^2}{qL(L-1)n}.$$

Proof. The proof is a simple application of (2.2). The second norm on the right-hand side can be expanded as

$$\begin{aligned} \left\| \sum_{i \in [L]} \varphi(c_i) \right\|_2^2 &= \sum_{i,j \in [L]} \langle \varphi(c_i), \varphi(c_j) \rangle \\ &\overset{(2.2)}{=} \sum_{i,j \in [L]} \Big((q-1)n - qn\delta(c_i, c_j) \Big) \\ &= L^2(q-1)n - 2qn \sum_{1 \leq i < j \leq L} \delta(c_i, c_j) \\ &= L^2(q-1)n - 2qn \binom{L}{2} \delta, \end{aligned}$$

and the bound follows. $\square$

Now we are ready to formulate our reduction from RIP-2 to average distance of codes.

[5] We stress that in this work, we crucially use the fact that the definition of RIP that we use is based on the Euclidean (ℓ_2) norm.

LEMMA 3.4. *Let $\mathcal{C} \subseteq [q]^n$ be a code and suppose $\varphi(\mathcal{C})/\sqrt{(q-1)n}$ satisfies RIP-2 of order L with constant $1/2$. Then, the average pairwise distance between every L codewords of $\mathcal{C}$ is at least $\big(1-\frac{1}{q}\big)\big(1-\frac{1}{2(L-1)}\big)$.*

Proof. Consider any set S of L codewords and the real vector $x \in \mathbb{R}^{|\mathcal{C}|}$ with entries in $\{0,1\}$ that is exactly supported on the positions indexed by the codewords in S. Obviously, $\|x\|_2^2 = L$. Thus, by the definition of RIP-2 (Definition 3.1), we know that, defining $M := \varphi(\mathcal{C})$,

$$\|Mx\|_2^2 \leq 3L(q-1)n/2. \tag{3.1}$$

Observe that $Mx = \sum_{i\in[L]} \varphi(c_i)$. Let δ be the average pairwise distance between codewords in S. By Lemma 3.3 we conclude that

$$\begin{aligned}\delta &= \frac{L^2(q-1)n - \Big\| \sum_{i\in[L]} \varphi(c_i) \Big\|_2^2}{2q\binom{L}{2}n} \\ &\overset{(3.1)}{\geq} \frac{(L^2-1.5L)(q-1)n}{qL(L-1)n} \\ &= \frac{q-1}{q}\Big(1-\frac{1}{2(L-1)}\Big). \qquad \square\end{aligned}$$

We remark that, for our applications, the exact choice of the RIP constant in the above result is arbitrary, as long as it remains an absolute constant (although the particular choice of the RIP constant would also affect the constants in the resulting bound on average pairwise distance). Contrary to applications in compressed sensing, for our application it also makes sense to have RIP-2 with constants larger than one, since the proof requires only the upper bound in Definition 3.1.

By combining Lemma 3.4 with the simplified Johnson bound of Theorem 3.2, we obtain the following corollary.

THEOREM 3.5. *Let $\mathcal{C} \subseteq [q]^n$ be a code and suppose $\varphi(\mathcal{C})/\sqrt{(q-1)n}$ satisfies RIP-2 of order L with constant $1/2$. Then $\mathcal{C}$ is $\Big(\big(1-\frac{1}{q}\big)\big(1-\sqrt{\frac{1.5}{L-1}}\big), L-1\Big)$-list decodable.*

Remark 3.6. Theorem 3.5 is a direct corollary of Lemma 3.4 and Theorem 3.2, which in turn follow from mathematically simple proofs and establish more general connections between the notion of average distance of codes, list decodability, and RIP. However, it is possible to directly prove Theorem 3.5 without establishing such independently interesting connections. One such proof is presented in Appendix B.

The matrix $\varphi(\mathcal{C})$ for a linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ has a special form. It is straightforward to observe that, when $q = 2$, the matrix is an incomplete Hadamard–Walsh transform matrix with $2^{\tilde{k}}$ columns, where $\tilde{k}$ is the dimension of the code. In general $\varphi(\mathcal{C})$ turns out to be related to a DFT matrix. Specifically, we have the following observation.

Observation 3.7. Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be an $[n, \tilde{k}]$ linear code with a generator matrix $G \in \mathbb{F}_q^{\tilde{k}\times n}$, and define $N := q^{\tilde{k}}$. Consider the matrix of *linear forms* $\mathsf{Lin} \in \mathbb{F}_q^{N\times N}$ with rows and columns indexed by elements of $\mathbb{F}_q^{\tilde{k}}$ and entries defined by

$$\mathsf{Lin}(x,y) := \langle x, y\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the finite-field inner product over $\mathbb{F}_q^{\tilde{k}}$. Let $T \subseteq \mathbb{F}_q^{\tilde{k}}$ be the multiset of columns of G. Then, $\varphi(\mathcal{C}) = \varphi(\mathsf{Lin}_T)$. (Recall from Definition 2.4 that the former simplex encoding $\varphi(\mathcal{C})$ is applied to the matrix enumerating the codewords of $\mathcal{C}$, while the latter, $\varphi(\mathsf{Lin}_T)$, is applied to the entries of a submatrix of Lin. Also recall from the notation section that Lin_T denotes the submatrix of Lin obtained by choosing all the rows of Lin indexed by the elements of the multiset T, with possible repetitions.)

When G is uniformly random, $\mathcal{C}$ becomes a random linear code and $\varphi(\mathcal{C})$ can be sampled by the following process. Arrange n uniformly random rows of Lin, sampled independently and with replacement, as rows of a matrix M. Then, replace each entry of M by its simplex encoding, seen as a column vector in $\mathbb{C}^{q-1}$. The resulting complex matrix is $\varphi(\mathcal{C})$.

The RIP-2 condition for random complex matrices arising from random linear codes is proved in Theorem 4.1. We now combine this theorem with the preceding results of this section to prove our main theorem on list decodability of random linear codes.

Theorem 3.8 (main). *Let q be a prime power, and let $\epsilon, \gamma > 0$ be constant parameters. Then for all large enough integers n, a random linear code $\mathcal{C} \subseteq \mathbb{F}_q^n$ of rate R for some*

$$R \gtrsim \frac{\epsilon^2}{\log(1/\gamma)\log^3(q/\epsilon)\log q}$$

is $((1-1/q)(1-\epsilon), O(1/\epsilon^2))$-list decodable with probability at least $1-\gamma$.

Proof. Let $\mathcal{C} \subseteq \mathbb{F}_q^n$ be a uniformly random linear code associated to a random $Rn \times n$ generator matrix G over $\mathbb{F}_q$ for a rate parameter $R \leq 1$ to be determined later. Consider the random matrix $M = \varphi(\mathcal{C}) = \varphi(\mathsf{Lin}_T)$ from Observation 3.7, where $|T| = n$. Recall that M is a $(q-1)n \times N$ complex matrix, where $N = q^{Rn}$. Let $L := 1 + \lceil 1.5/\epsilon^2 \rceil = \Theta(1/\epsilon^2)$. By Theorem 4.1, for large enough N (thus large enough n) and with probability $1-\gamma$, the matrix $M/\sqrt{(q-1)n}$ satisfies RIP-2 of order L with constant $1/2$, for some choice of $|T|$ bounded by

$$n = |T| \lesssim \log(1/\gamma) L \log(N) \log^3(qL). \tag{3.2}$$

Suppose n is large enough and satisfies (3.2) so that the RIP-2 condition holds. By Theorem 3.5, this ensures that the code $\mathcal{C}$ is $((1-1/q)(1-\epsilon), O(1/\epsilon^2))$-list decodable with probability at least $1-\gamma$.

It remains to verify the bound on the rate of $\mathcal{C}$. We observe that whenever the RIP-2 condition is satisfied, G must have rank exactly Rn, since otherwise there would be distinct vectors $x, x' \in \mathbb{F}_q^{Rn}$ such that $xG = x'G$. Thus in that case, the columns of M corresponding to x and x' become identical, implying that M cannot satisfy RIP-2 of any nontrivial order. Thus we can assume that the rate of $\mathcal{C}$ is indeed equal to R. Now we have

$$\begin{aligned} R &= \log_q |\mathcal{C}|/n = \log N/(n \log q) \\ &\overset{(3.2)}{\gtrsim} \frac{\log N}{\log(1/\gamma) L \log(N) \log^3(qL) \log q}. \end{aligned}$$

Substituting $L = \Theta(1/\epsilon^2)$ into the above expression yields the desired bound. □

4. RIP of DFT-based matrices.

In this section, we prove RIP-2 for random incomplete DFT matrices. Namely, we prove the following theorem.

Theorem 4.1. *Let T be a random multiset of rows of Lin, where $|T|$ is fixed and each element of T is chosen uniformly at random and independently with replacement.*

Then, for every $\delta, \gamma > 0$, and assuming $N \geq N_0(\delta, \gamma)$, with probability at least $1-\gamma$ the matrix $\varphi(\mathsf{Lin}_T)/\sqrt{(q-1)|T|}$ (with $(q-1)|T|$ rows) satisfies RIP-2 of order k with constant δ for a choice of $|T|$ satisfying

$$|T| \lesssim \frac{\log(1/\gamma)}{\delta^2} k \log(N) \log^3(qk). \tag{4.1}$$

The proof extends and closely follows the original proof of Rudelson and Vershynin [27]. However, we modify the proof at a crucial point to obtain a strict improvement over their original analysis, which is necessary for our list decoding application. We present our improved analysis in this section.

Proof of Theorem 4.1. Let $M := \varphi(\mathsf{Lin}_T)$. Each row of M is indexed by an element of T and some $\alpha \in \mathbb{F}_q^*$, where in the definition of simplex encoding (Definition 2.4), we identify $\mathbb{F}_q^*$ with $[q-1]$ in a fixed but arbitrary way. Recall that $T \subseteq \mathbb{F}_q^{\tilde{k}}$, where $N = q^{\tilde{k}}$. Denote the row corresponding to $t \in T$ and $\alpha \in \mathbb{F}_q^*$ by $M_{t,\alpha}$, and moreover denote the set of k-sparse unit vectors in $\mathbb{C}^N$ by $\mathcal{B}_2^{k,N}$.

In order to show that $M/\sqrt{(q-1)|T|}$ satisfies RIP of order k, we need to verify that for any $x = (x_1, \ldots, x_N) \in \mathcal{B}_2^{k,N}$,

$$|T|(q-1)(1-\delta) \leq \|Mx\|_2^2 \leq |T|(q-1)(1+\delta). \tag{4.2}$$

In light of Proposition A.1, without loss of generality we can assume that x is real-valued (since the inner product between any pair of columns of M is real-valued).

For $i \in \mathbb{F}_q^n$, denote the ith column of M by M^i. For $x = (x_1, \ldots, x_N) \in \mathcal{B}_2^{k,N}$, define the random variable

$$\begin{aligned} \Delta_x &:= \|Mx\|_2^2 - |T|(q-1) \\ &= \sum_{\substack{(i,j) \in \mathsf{supp}(x) \\ i \neq j}} x_i x_j \langle M^i, M^j \rangle, \end{aligned} \tag{4.3}$$

where the second equality holds since each column of M has ℓ_2 norm $\sqrt{(q-1)|T|}$ and $\|x\|_2 = 1$. Thus, the RIP condition (4.2) is equivalent to

$$\Delta := \sup_{x \in \mathcal{B}_2^{k,N}} |\Delta_x| \leq \delta |T|(q-1). \tag{4.4}$$

Recall that Δ is a random variable depending on the randomness in T. The proof of the RIP condition involves two steps: first, bounding Δ in expectation, and second, a tail bound. The first step is proved, in detail, in the following lemma.

LEMMA 4.2. *Let $\delta' > 0$ be a real parameter. Then, $\mathbb{E}[\Delta] \leq \delta' |T|(q-1)$ for a choice of $|T|$ bounded as follows:*

$$|T| \lesssim k \log(N) \log^3(qk)/\delta'^2.$$

Proof. We begin by observing that the columns of M are orthogonal in expectation; i.e., for any $i, j \in \mathbb{F}_q^n$, we have

$$\mathbb{E}_T \langle M^i, M^j \rangle = \begin{cases} |T|(q-1), & i = j, \\ 0, & i \neq j. \end{cases}$$

This follows from (2.2) and the fact that the expected relative Hamming distance between the columns of Lin corresponding to i and j, when $i \neq j$, is exactly $1-1/q$. It follows that for every $x \in \mathcal{B}_2^{k,N}$, $\mathbb{E}[\Delta_x] = 0$, namely, the stochastic process $\{\Delta_x\}_{x\in\mathcal{B}_2^{k,N}}$ is centered.

Recall that we wish to estimate

$$\begin{aligned} \mathcal{E} &:= \mathbb{E}_T \Delta \\ &= \mathbb{E}_T \sup_{x\in\mathcal{B}_2^{k,N}} \left| \sum_{t\in T} \sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x\rangle^2 - |T|(q-1) \right|. \end{aligned} \tag{4.5}$$

Suppose the chosen multiset of the rows of Lin is written as a random sequence $T = (t_1, t_2, \ldots, t_{|T|})$. The random variables $\sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t_i,\alpha}, x\rangle^2$, for different values of i, are independent. Therefore, we can use the standard symmetrization technique on summation of independent random variables in a stochastic process (Proposition A.2) and conclude from (4.5) that

$$\mathcal{E} \lesssim \mathcal{E}_1 := \mathbb{E}_T \mathbb{E}_{\mathcal{G}} \sup_{x\in\mathcal{B}_2^{k,N}} \left(\sum_{t\in T} g_t \sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x\rangle^2 \right), \tag{4.6}$$

where $\mathcal{G} := (g_t)_{t\in T}$ is a sequence of independent standard Gaussian random variables. Denote the term inside $\mathbb{E}_T$ in (4.6) by $\mathcal{E}_T$; namely,

$$\mathcal{E}_T := \mathbb{E}_{\mathcal{G}} \sup_{x\in\mathcal{B}_2^{k,N}} \left(\sum_{t\in T} g_t \sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x\rangle^2 \right).$$

Now we observe that, for any fixed T, the quantity $\mathcal{E}_T$ defines the supremum of a Gaussian process. The Gaussian process $\{G_x\}_{x\in\mathcal{B}_2^{k,N}}$ induces a pseudometric $\|\cdot\|_X$ on $\mathcal{B}_2^{k,N}$ (and, more generally, $\mathbb{C}^N$), where for $x, x' \in \mathcal{B}_2^{k,N}$, the (squared) distance is given by

$$\begin{aligned} \|x-x'\|_X^2 &:= \mathbb{E}_G |G_x - G_{x'}|^2 \\ &= \sum_{t\in T} \left(\sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x\rangle^2 - \sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x'\rangle^2 \right)^2 \\ &= \sum_{t\in T} \left(\sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x+x'\rangle \langle M_{t,\alpha}, x-x'\rangle \right)^2. \end{aligned} \tag{4.7}$$

By Cauchy–Schwarz, (4.7) can be bounded as

$$\|x-x'\|_X^2 \le \sum_{t\in T} \left(\sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x+x'\rangle^2 \right) \left(\sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x-x'\rangle^2 \right) \tag{4.8}$$

$$\le \sum_{t\in T} \sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x+x'\rangle^2 \max_{t\in T} \left(\sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, x-x'\rangle^2 \right). \tag{4.9}$$

Here is where our analysis differs from [27]. When $q = 2$, (4.9) is exactly how the Gaussian metric is bounded in [27]. We obtain our improvement by bounding the metric in a different way. Specifically, let $\eta \in (0, 1]$ be a positive real parameter to be determined later and let $r := 1 + \eta$ and $s := 1 + 1/\eta$ such that $1/r + 1/s = 1$. We assume that η is so that s becomes an integer. We use Hölder's inequality with parameters r and s along with (4.8) to bound the metric as follows:

$$(4.10)\quad \|x - x'\|_X \\ \le \left(\sum_{t\in T}\left(\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x + x'\rangle^2\right)^r\right)^{1/2r}\left(\sum_{t\in T}\Big(\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x - x'\rangle^2\Big)^s\right)^{1/2s}.$$

Since $\|x\|_2 = 1$, x is k-sparse, and $|M_{t,\alpha}| = 1$ for all choices of (t, α), Cauchy–Schwarz implies that $\langle M_{t,\alpha}, x\rangle^2 \le k$ and thus, using the triangle inequality, we know that

$$\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x + x'\rangle^2 \le 4qk.$$

Therefore, for every $t \in T$, seeing that $r = 1 + \eta$, we have

$$\left(\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x + x'\rangle^2\right)^r \le (4qk)^\eta \sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x + x'\rangle^2,$$

which, applied to the bound (4.10) on the metric, yields

$$(4.11)\quad \|x - x'\|_X \\ \le (4qk)^{\eta/2r}\underbrace{\left(\sum_{t\in T}\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x + x'\rangle^2\right)}_{\mathcal{E}_2}^{1/2r}\left(\sum_{t\in T}\left(\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x - x'\rangle^2\right)^s\right)^{1/2s}.$$

Now,

$$(4.12)\qquad \mathcal{E}_2 \le 2\left(\sum_{t\in T}\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x\rangle^2 + \sum_{t\in T}\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x'\rangle^2\right) \le 4\mathcal{E}'_T,$$

where we have defined

$$(4.13)\qquad \mathcal{E}'_T := \sup_{x\in\mathcal{B}_2^{k,N}}\sum_{t\in T}\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x\rangle^2.$$

Observe that, by the triangle inequality,

$$(4.14)\qquad \mathcal{E}'_T \le \sup_{x\in\mathcal{B}_2^{k,N}}\left|\sum_{t\in T}\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x\rangle^2 - |T|(q-1)\right| + |T|(q-1).$$

Plugging (4.13) back in (4.11), we so far have

$$\|x-x'\|_X \le 4(4qk)^{\eta/2r}{\mathcal{E}'_T}^{1/2r}\left(\sum_{t\in T}\left(\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x-x'\rangle^2\right)^s\right)^{1/2s}. \tag{4.15}$$

For a real parameter $u>0$, define $N_X(u)$ as the minimum number of spheres of radius u required to cover $\mathcal{B}_2^{k,N}$ with respect to the metric $\|\cdot\|_X$. We can now apply Dudley's theorem on supremum of Gaussian processes (cf. [25, Theorem 11.17]) and deduce that

$$\mathcal{E}_T \lesssim \int_0^\infty \sqrt{\log N_X(u)}du. \tag{4.16}$$

In order to make the metric $\|\cdot\|_X$ easier to work with, we define a related metric $\|\cdot\|_{X'}$ on $\mathcal{B}_2^{k,N}$, according to the right-hand side of (4.15), as follows:

$$\|x-x'\|_{X'}^{2s} := \sum_{t\in T}\left(\sum_{\alpha\in\mathbb{F}_q^*}\langle M_{t,\alpha}, x-x'\rangle^2\right)^s. \tag{4.17}$$

Let K denote the diameter of $\mathcal{B}_2^{k,N}$ under the metric $\|\cdot\|_{X'}$. Trivially, $K\le 2|T|^{1/2s}\sqrt{qk}$. By (4.15), we know that

$$\|x-x'\|_X \le 4(4qk)^{\eta/2r}{\mathcal{E}'_T}^{1/2r}\|x-x'\|_{X'}. \tag{4.18}$$

Define $N_{X'}(u)$ similar to $N_X(u)$, but with respect to the new metric X'. The preceding upper bound (4.18) thus implies that

$$N_X(u) \le N_{X'}(u/(4(4qk)^{\eta/2r}{\mathcal{E}'_T}^{1/2r})). \tag{4.19}$$

Now, using this bound in (4.16) and after a change of variables, we have

$$\mathcal{E}_T \lesssim (4qk)^{\eta/2r}{\mathcal{E}'_T}^{1/2r}\int_0^\infty \sqrt{\log N_{X'}(u)}du. \tag{4.20}$$

Now we take an expectation over T. Note that (4.14) combined with (4.5) implies

$$\mathbb{E}_T\mathcal{E}'_T \le \mathcal{E} + |T|(q-1). \tag{4.21}$$

Using (4.16), we get

$$\begin{aligned}
\mathcal{E}^{2r} &\overset{(4.6)}{\lesssim} \mathcal{E}_1^{2r} = (\mathbb{E}_T\mathcal{E}_T)^{2r} \le \mathbb{E}_T\mathcal{E}_T^{2r}\\
&\lesssim (4qk)^\eta\mathbb{E}_T\left((\mathcal{E}'_T)^{1/2r}\int_0^\infty\sqrt{\log N_{X'}(u)}du\right)^{2r}\\
&\le (4qk)^\eta(\mathbb{E}_T\mathcal{E}'_T)\max_T\left(\int_0^\infty\sqrt{\log N_{X'}(u)}du\right)^{2r}\\
&\overset{(4.21)}{\le} (4qk)^\eta(\mathcal{E}+|T|(q-1))\max_T\left(\int_0^\infty\sqrt{\log N_{X'}(u)}du\right)^{2r}.
\end{aligned}$$

Define

$$\bar{\mathcal{E}} := \mathcal{E} \cdot \left(\frac{\mathcal{E}}{\mathcal{E} + |T|(q-1)} \right)^{1/(1+2\eta)}. \tag{4.22}$$

Therefore, recalling that $r = 1 + \eta$, the above inequality simplifies to

$$\bar{\mathcal{E}} \lesssim (4qk)^{\eta} \max_T \left(\int_0^K \sqrt{\log N_{X'}(u)} du \right)^{1+1/(1+2\eta)}, \tag{4.23}$$

where we have replaced the upper limit of integration by the diameter of $\mathcal{B}_2^{k,N}$ under the metric $\|\cdot\|_{X'}$. (Obviously, $N_{X'}(u) = 1$ for all $u \geq K$.)

Now we estimate $N_{X'}(u)$ in two ways. The first estimate is the simple volumetric estimate (cf. [27]) that gives

$$\log N_{X'}(u) \lesssim k \log(N/k) + k \log(1 + 2K/u). \tag{4.24}$$

This estimate is useful when u is small. For larger values of u, we use a different estimate as follows.

CLAIM 4.3. $\log N_{X'}(u) \lesssim |T|^{1/s} (\log N) qks/u^2$.

Proof. We use the method used in [27] (originally attributed to Maurey; cf. [9, section 1]) and empirically estimate any fixed real vector $x = (x_1, \ldots, x_N) \in \mathcal{B}_2^{k,N}$ by an m-sparse random vector Z for sufficiently large m. The vector Z is an average

$$Z := \frac{\sqrt{k}}{m} \sum_{i=1}^{m} Z_i, \tag{4.25}$$

where each Z_i is a 1-sparse vector in $\mathbb{C}^N$ and $\mathbb{E}[Z_i] = x/\sqrt{k}$. The Z_i are independent and identically distributed.

The way each Z_i is sampled is as follows. Let $x' := x/\sqrt{k}$ so that $\|x'\|_1 = \frac{\|x\|_1}{\sqrt{k}} \leq 1$. With probability $1 - \|x'\|$, we set $Z_i := 0$. With the remaining probability, Z_i is sampled by picking a random $j \in \mathsf{supp}(x)$ according to the probabilities defined by absolute values of the entries of x' and setting $Z_i = \mathrm{sgn}(x'_j) e_j$, where e_j is the jth standard basis vector.[6] This ensures that $\mathbb{E}[Z_i] = x'$. Thus, by linearity of expectation, it is clear that $\mathbb{E}[Z] = x$. Now, consider

$$\mathcal{E}_3 := \mathbb{E}\|Z - x\|_{X'}.$$

If we pick m large enough to ensure that $\mathcal{E}_3 \leq u$, regardless of the initial choice of x, then we can conclude that for every x, there exists a Z of the form (4.25) that is at distance at most u from x (since there is always some fixing of the randomness that attains the expectation). In particular, the set of balls centered at all possible realizations of Z would cover $\mathcal{B}_2^{k,N}$. Since the number of possible choices of Z of the form (4.25) is at most $(2N+1)^m$, we have

$$\log N_{X'}(u) \lesssim m \log N. \tag{4.26}$$

[6] Note that since we have assumed x is a real vector, $\mathrm{sgn}(\cdot)$ is always well defined.

In order to estimate the number of independent samples m, we use symmetrization again to estimate the deviation of Z from its expectation x. Namely, since the Z_i are independent, by the symmetrization technique stated in Proposition A.2 we have

$$\mathcal{E}_3 \lesssim \frac{\sqrt{k}}{m} \cdot \mathbb{E} \left\| \sum_{i=1}^{m} \epsilon_i Z_i \right\|_{X'}, \tag{4.27}$$

where $(\epsilon_i)_{i\in[m]}$ is a sequence of independent Rademacher random variables in $\{-1,+1\}$. Now, consider

$$\begin{aligned} \mathcal{E}_4 &:= \mathbb{E} \left\| \sum_{i=1}^{m} \epsilon_i Z_i \right\|_{X'}^{2s} \\ &= \mathbb{E} \sum_{t\in T} \left(\sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, \sum_{i=1}^{m} \epsilon_i Z_i \rangle^2 \right)^s \\ &= \sum_{t\in T} \mathbb{E} \left(\sum_{\alpha\in\mathbb{F}_q^*} \left(\sum_{i=1}^{m} \epsilon_i \langle M_{t,\alpha}, Z_i \rangle \right)^2 \right)^s \\ &= \sum_{t\in T} \mathbb{E} \left(\sum_{i,j=1}^{m} \epsilon_i\epsilon_j \sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, Z_i \rangle \langle M_{t,\alpha}, Z_j \rangle^* \right)^s . \end{aligned} \tag{4.28}$$

Since the entries of the matrix M are bounded in magnitude by 1, we have

$$\left| \sum_{\alpha\in\mathbb{F}_q^*} \langle M_{t,\alpha}, Z_i \rangle \langle M_{t,\alpha}, Z_j \rangle^* \right| \leq q.$$

Using this bound and Proposition A.3, (4.28) can be simplified as

$$\mathcal{E}_4 = \mathbb{E} \left\| \sum_{i=1}^{m} \epsilon_i Z_i \right\|_{X'}^{2s} \leq |T|(4qms)^s,$$

and combined with (4.27), and using Jensen's inequality,

$$\mathcal{E}_3 \lesssim |T|^{1/2s} \sqrt{4qks/m}.$$

Therefore, we can ensure that $\mathcal{E}_3 \leq u$, as desired, for some large enough choice of m, specifically, for some $m \lesssim |T|^{1/s} qks/u^2$. Now from (4.26), we get

$$\log N_{X'}(u) \lesssim |T|^{1/s} (\log N) qks/u^2. \tag{4.29}$$

Claim 4.3 is now proved. $\square$

Now we continue the proof of Lemma 4.2. Break the integration in (4.23) into two intervals. Consider

$$\mathcal{E}_5 := \underbrace{\int_0^A \sqrt{\log N_{X'}(u)}du}_{\mathcal{E}_6} + \underbrace{\int_A^K \sqrt{\log N_{X'}(u)}du}_{\mathcal{E}_7},$$

where $A := K/\sqrt{qk}$. We claim the following bound on $\mathcal{E}_5$.

CLAIM 4.4. $\mathcal{E}_5 \lesssim |T|^{1/2s}\sqrt{(\log N)qks}\log(qk)$.

Proof. First, we use (4.24) to bound $\mathcal{E}_6$ as follows:

$$\mathcal{E}_6 \lesssim A\sqrt{k\log(N/k)} + \sqrt{k}\int_0^A \sqrt{\ln(1+2K/u)}du. \tag{4.30}$$

Observe that $2K/u \geq 1$, so $1 + 2K/u \leq 4K/u$. Thus,

$$\begin{aligned}\int_0^A \sqrt{\ln(1+2K/u)}\,du &\leq \int_0^A \sqrt{\ln(4K/u)}\,du \\ &= 2K\int_0^{A/2K} \sqrt{\ln(2/u)}\,du \\ &= 2K\left(\frac{A}{2K}\sqrt{\ln(4K/A)} + \sqrt{\pi}\left(1 - \operatorname{erf}\left(\sqrt{\ln(4K/A)}\right)\right)\right) \\ &= A\sqrt{\ln(4K/A)} + 2\sqrt{\pi}K\operatorname{erfc}\left(\sqrt{\ln(4K/A)}\right), \qquad (4.31)\end{aligned}$$

where $\operatorname{erf}(\cdot)$ is the Gaussian error function $\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$, and $\operatorname{erfc}(x) := 1 - \operatorname{erf}(x)$, and we have used the integral identity

$$\int \sqrt{\ln(1/x)}dx = -\frac{\sqrt{\pi}}{2}\operatorname{erf}\left(\sqrt{\ln(1/x)}\right) + x\sqrt{\ln(1/x)} + C$$

that can be verified by taking derivatives of both sides. Let us use the upper bound

$$(\text{for all } x > 0)\quad \operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt \leq \frac{2}{\sqrt{\pi}}\int_x^\infty \frac{t}{x}e^{-t^2}dt = \frac{1}{\sqrt{\pi}}\cdot\frac{e^{-x^2}}{x},$$

and plug it into (4.31) to obtain

$$\begin{aligned}\int_0^A \sqrt{\ln(1+2K/u)}\,du &\leq A\sqrt{\ln(4K/A)} + 2\sqrt{\pi}K\left(\frac{1}{\sqrt{\pi}}\cdot\frac{A}{4K}\cdot\frac{1}{\sqrt{\ln(4K/A)}}\right) \\ &= A\sqrt{\ln(4K/A)} + \frac{A}{2\sqrt{\ln(4K/A)}} \\ &\lesssim A\sqrt{\log(qk)} \lesssim |T|^{1/2s}\sqrt{\log(qk)}),\end{aligned}$$

where the last inequality holds since $A = K/\sqrt{qk} \lesssim |T|^{1/2s}$. Therefore, by (4.30) we get

$$\mathcal{E}_6 \lesssim |T|^{1/2s}\sqrt{k}(\sqrt{\log N} + \sqrt{\log(qk)}). \tag{4.32}$$

On the other hand, we use Claim 4.3 to bound $\mathcal{E}_7$:

$$\begin{aligned} \mathcal{E}_7 &\lesssim \sqrt{|T|^{1/s}(\log N)qks}\int_A^K du/u \\ &\lesssim |T|^{1/2s}\sqrt{(\log N)qks}\log(qk). \end{aligned} \tag{4.33}$$

Combining (4.32) and (4.33), we conclude that for every fixed T,

$$\mathcal{E}_5 = \mathcal{E}_6 + \mathcal{E}_7 \lesssim |T|^{1/2s}\sqrt{(\log N)qks}\log(qk).$$

Claim 4.4 is now proved. □

By combining Claim 4.4 and (4.23), we have

$$\begin{aligned} \bar{\mathcal{E}} &\lesssim (4qk)^\eta \max_T \mathcal{E}_5^{1+1/(1+2\eta)} \\ &\lesssim (4qk)^\eta \Big(|T|^{1/2s}\sqrt{(\log N)qks}\log(qk)\Big)^{1+1/(1+2\eta)} \\ &= (4qk)^\eta |T|^{\eta/(1+2\eta)}\Big(\sqrt{(\log N)qks}\log(qk)\Big)^{1+1/(1+2\eta)}. \end{aligned} \tag{4.34}$$

By Proposition A.4 (setting $a := \mathcal{E}/(|T|(q-1))$ and $\mu := 2\eta$), and recalling the definition (4.22) of $\bar{\mathcal{E}}$, in order to ensure that $\mathcal{E} \le \delta'(q-1)|T|$, it suffices to have

$$\bar{\mathcal{E}} \le \delta'^{\frac{2(1+\eta)}{1+2\eta}}|T|(q-1)/4. \tag{4.35}$$

Using (4.34), and after simple manipulations, (4.35) can be ensured for some

$$|T| \lesssim \frac{(4qk)^{2\eta}}{\eta}k(\log N)\log^2(qk)/\delta'^2.$$

This expression is minimized for some $\eta = 1/\Theta(\log(qk))$, which gives

$$|T| \lesssim k(\log N)\log^3(qk)/\delta'^2.$$

This concludes the proof of Lemma 4.2. □

Now we turn to the tail bound on the random variable Δ and estimate the appropriate size of T required to ensure that $\Pr[\Delta > \delta|T|(q-1)] \le \gamma$. We observe that the tail bound proved in [27] uses the bound on $\mathbb{E}[\Delta]$ as a black box. In particular, the following lemma, for $q = 2$, is implicit in the proof of Theorem 3.9 in [27]. The extension to arbitrary alphabet size q and our slightly different subsampling process is straightforward. However, for completeness, we include a detailed proof of Lemma 4.5 in Appendix C.

Lemma 4.5 (see [27, implicit]). *Suppose that for some $\delta' > 0$, $\mathbb{E}[\Delta] \le \delta'|T|(q-1)$. Then, there are absolute constants c_1, c_2, c_3 such that for every $\lambda \ge 1$,*

$$\Pr[\Delta > (c_1 + c_2\lambda)\delta'|T|(q-1)] \le 6\exp(-\lambda^2),$$

provided that

$$|T|/k \geq c_3\lambda/\delta'. \tag{4.36}$$

Proof. See Appendix C. □

Now it suffices to instantiate the above lemma with $\lambda := \sqrt{\ln(6/\gamma)}$ and $\delta' := \delta/(c_1 + c_2\lambda) = \delta/\Theta(\sqrt{\ln(6/\gamma)})$ and use the resulting value of δ' in Lemma 4.2. Since Lemma 4.2 ensures that $|T|/k = \Omega(\log N)$, we can take N large enough (depending on δ, γ) so that (4.36) is satisfied. This completes the proof of Theorem 4.1. □

The proof of Theorem 4.1 does not use any property of the DFT-based matrix other than orthogonality and boundedness of the entries. However, for syntactical reasons, that is, the way the matrix is defined for $q > 2$, we have presented the theorem and its proof for the special case of the DFT-based matrices. The proof goes through with no technical changes for any orthogonal matrix with bounded entries (as is the case for the original proof of [27]). In particular, we remark that the following variation of Theorem 4.1 also holds.

THEOREM 4.6. *Let $A \in \mathbb{C}^{N\times N}$ be any orthonormal matrix with entries bounded by $O(1/\sqrt{N})$. Let T be a random multiset of rows of A, where $|T|$ is fixed and each element of T is chosen uniformly at random and independently with replacement. Then, for every $\delta, \gamma > 0$, and assuming $N \geq N_0(\delta, \gamma)$, with probability at least $1-\gamma$ the matrix $(\sqrt{N/|T|})A_T$ satisfies RIP-2 of order k with constant δ for a choice of $|T|$ satisfying*

$$|T| \lesssim \frac{\log(1/\gamma)}{\delta^2} k(\log N)\log^3 k.$$

We also note that the subsampling procedure required by Theorem 4.1 is slightly different from the one originally used by [27]. In our setting, we appropriately fix the target number of row (i.e., $|T|$) first and then draw as many uniform and independent samples of the rows of the original Fourier matrix as needed (with replacement). On the other hand, [27] samples the RIP matrix by starting from the original $N \times N$ Fourier matrix and then removing each row independently with a certain probability. This probability is carefully chosen so that the expected number of remaining rows in the end of the process is sufficiently large. Our modified sampling is well suited for our coding-theoretic applications and offers the additional advantage of always returning a matrix with the exact desired number of rows. However, we point out that since Theorem 4.1 is based on the original ideas of [27], it can be verified to hold with respect to either of the two subsampling procedures.

Appendix A. Useful tools. The original definition of RIP-2 given in Definition 3.1 considers all complex vectors $x \in \mathbb{C}^n$. Below we show that it suffices to satisfy the property only for real-valued vectors x.

PROPOSITION A.1. *Let $M \in \mathbb{C}^{m\times N}$ be a complex matrix such that $M^\dagger M \in \mathbb{R}^{N\times N}$ and for any k-sparse vector $x \in \mathbb{R}^N$, we have*

$$(1-\delta)\|x\|_2^2 \leq \|Mx\|_2^2 \leq (1+\delta)\|x\|_2^2.$$

Then, M satisfies RIP-2 of order k with constant δ.

Proof. Let $x = a + \mathbf{i}b$, for some $a, b \in \mathbb{R}^N$, be any complex vector. We have $\|x\|_2^2 = \|a\|_2^2 + \|b\|_2^2$, and

$$\begin{aligned}
\left|\|Mx\|_2^2 - \|x\|_2^2\right| &= \left|x^\dagger M^\dagger M x - \|x\|_2^2\right| \\
&= \left|(a^\dagger - \mathbf{i}b^\dagger)M^\dagger M(a+\mathbf{i}b) - \|x\|_2^2\right| \\
&= \left|a^\dagger M^\dagger M a + b^\dagger M^\dagger M b + \mathbf{i}(a^\dagger M^\dagger M b - b^\dagger M^\dagger M a) - \|x\|_2^2\right| \\
&\overset{(\star)}{=} \left|a^\dagger M^\dagger M a + b^\dagger M^\dagger M b - \|x\|_2^2\right| \\
&= \left|a^\dagger M^\dagger M a - \|a\|_2^2 + b^\dagger M^\dagger M b - \|b\|_2^2\right| \\
&\overset{(\star\star)}{\leq} \delta\|a\|_2^2 + \delta\|b\|_2^2 \\
&= \delta\|x\|_2^2,
\end{aligned}$$

where $(\star)$ is due to the assumption that $M^\dagger M$ is real, which implies that $a^\dagger M^\dagger M b$ and $b^\dagger M^\dagger M a$ are conjugate real numbers (and thus equal), and $(\star\star)$ is from the assumption that the RIP-2 condition is satisfied by M for real-valued vectors and the triangle inequality. □

As a technical tool, we use the standard symmetrization technique summarized in the following proposition for bounding deviation of summation of independent random variables from the expectation. The proof is a simple convexity argument (see, e.g., [25, Lemma 6.3] and [31, Lemma 5.70]).

Proposition A.2. *Let $(X_i)_{i\in[m]}$ be a finite sequence of independent random variables in a Banach space, and let $(\epsilon_i)_{i\in[m]}$ and $(g_i)_{i\in[m]}$ be sequences of independent Rademacher (i.e., each uniformly random in $\{-1,+1\}$) and standard Gaussian random variables, respectively. Then,*

$$\mathbb{E}\left\|\sum_{i\in[m]}(X_i - \mathbb{E}[X_i])\right\| \lesssim \mathbb{E}\left\|\sum_{i\in[m]}\epsilon_i X_i\right\| \lesssim \mathbb{E}\left\|\sum_{i\in[m]} g_i X_i\right\|.$$

More generally, for a stochastic process $(X_i^{(\tau)})_{i\in[m],\tau\in\mathcal{T}}$ where $\mathcal{T}$ is an index set,

$$\mathbb{E}\sup_{\tau\in\mathcal{T}}\left\|\sum_{i\in[m]}\left(X_i^{(\tau)} - \mathbb{E}[X_i^{(\tau)}]\right)\right\| \lesssim \mathbb{E}\sup_{\tau\in\mathcal{T}}\left\|\sum_{i\in[m]}\epsilon_i X_i^{(\tau)}\right\| \lesssim \mathbb{E}\sup_{\tau\in\mathcal{T}}\left\|\sum_{i\in[m]} g_i X_i^{(\tau)}\right\|.$$

The following bound is used in the proof of Claim 4.3, a part of the proof of Lemma 4.2.

Proposition A.3. *Let $(\epsilon_i)_{i\in[m]}$ be a sequence of independent Rademacher random variables, and let $(a_{ij})_{i,j\in[m]}$ be a sequence of complex coefficients with magnitude bounded by K. Then,*

$$\left|\mathbb{E}\left(\sum_{i,j\in[m]} a_{ij}\epsilon_i\epsilon_j\right)^s\right| \leq (4Kms)^s.$$

Proof. By linearity of expectation, we can expand the moment as follows:

$$\mathbb{E}\left(\sum_{i,j\in[m]} a_{ij}\epsilon_i\epsilon_j\right)^s = \sum_{\substack{(i_1,\dots i_s)\in[m]^s \\ (j_1,\dots j_s)\in[m]^s}} \left(a_{i_1 j_1}\cdots a_{i_s j_s}\mathbb{E}\Big[\epsilon_{i_1}\cdots\epsilon_{i_s}\epsilon_{j_1}\cdots\epsilon_{j_s}\Big]\right).$$

Observe that $\mathbb{E}[\epsilon_{i_1}\cdots\epsilon_{i_s}\epsilon_{j_1}\cdots\epsilon_{j_s}]$ is equal to 1 whenever all integers in the sequence

$$(i_1,\ldots,i_s,j_1,\ldots,j_s)$$

appear an even number of times. Otherwise the expectation is zero. Denote by $S\subseteq[m]^{2s}$ the set of sequences $(i_1,\ldots,i_s,j_1,\ldots,j_s)$ that make the expectation nonzero. Then,

$$\left|\mathbb{E}\left(\sum_{i,j\in[m]} a_{ij}\epsilon_i\epsilon_j\right)^s\right| = \left|\sum_{(i_1,\ldots i_s,j_1,\ldots j_s)\in S} a_{i_1j_1}\cdots a_{i_sj_s}\right| \le K^s|S|.$$

One way to generate a sequence $\sigma\in S$ is as follows. Pick s coordinate positions of σ out of the $2s$ available positions, fill out each position by an integer in $[m]$, duplicate each integer at an available unpicked slot (in some fixed order), and finally permute the s positions of σ that were not originally picked. Obviously, this procedure can generate every sequence in S (although some sequences may be generated in many ways). The number of combinations that the combinatorial procedure can produce is bounded by $\binom{2s}{s}m^s(s!)\le(4ms)^s$. Therefore, $|S|\le(4ms)^s$ and the bound follows. □

We have used the following technical statement in the proof of Lemma 4.2.

PROPOSITION A.4. *Suppose for real numbers $a>0$, $\mu\in[0,1]$, $\delta\in(0,1]$, we have*

$$a\cdot\left(\frac{a}{1+a}\right)^{\frac{1}{1+\mu}} \le \frac{\delta^{\frac{2+\mu}{1+\mu}}}{4}.$$

Then, $a\le\delta$.

Proof. Let $\delta' := \delta^{\frac{2+\mu}{1+\mu}}/4^{\frac{1}{1+\mu}} \ge \delta^{\frac{2+\mu}{1+\mu}}/4$. From the assumption, we have

$$a\cdot\left(\frac{a}{1+a}\right)^{\frac{1}{1+\mu}} \le \delta' \Rightarrow a^{2+\mu}\le\delta^{2+\mu}(1+a)/4. \tag{A.1}$$

Consider the function

$$f(a) := a^{2+\mu}-\delta^{2+\mu}a/4-\delta^{2+\mu}/4.$$

The proof is complete if we show that, for every $a>0$, the assumption $f(a)\le 0$ implies $a\le\delta$, or equivalently, $a>\delta\Rightarrow f(a)>0$. Note that $f(0)<0$, and $f''(a)>0$ for all $a>0$. The function f attains a negative value at zero and is convex at all points $a>0$. Therefore, it suffices to show that $f(\delta)>0$. Now,

$$f(\delta)=\delta^{2+\mu}-\delta^{3+\mu}/4-\delta^{2+\mu}/4 \ge (3\delta^{2+\mu}-\delta^{3+\mu})/4.$$

Since $\delta\le 1$, the last expression is positive, and the claim follows. □

Appendix B. Direct proof of Theorem 3.5.

Let $\epsilon := \sqrt{\frac{1.5}{L-1}}$ and $M := \varphi(\mathcal{C})/\sqrt{(q-1)n}$. Let $S\subseteq\mathcal{C}$ be a set of L codewords, and suppose for the sake of contradiction that there is a vector $w\in[q]^n$ that is close in Hamming distance to all the L codewords in S. Namely, for each $c\in S$ we have

$$\delta(w,c) < \left(1-\frac{1}{q}\right)(1-\epsilon). \tag{B.1}$$

Let M' be the $(q-1)n \times L$ submatrix of M formed by removing all the columns of M corresponding to codewords of $\mathcal{C}$ outside the set S, and define $v := \varphi(w)/\sqrt{(q-1)n}$, considered as a row vector. RIP implies that for every non-zero vector $x \in \mathbb{R}^L$,

$$\frac{\|M'x\|_2^2}{\|x\|_2^2} \leq 3/2.$$

That is, if σ denotes the largest singular value of M', we have $\sigma^2 \leq 3/2$. Let $u := vM'$. From (B.1) combined with (2.2), we know that all the entries of u are greater than ϵ. Thus, $\|u\|_2^2 > \epsilon^2 L > 3/2$. On the other hand, $\|v\|_2 = 1$. This means that $\|vM'\|_2^2/\|v\|_2^2 > 3/2$, contradicting the bound on σ (maximum singular value of M').

Appendix C. Proof of Lemma 4.5. We closely follow the proof of Theorem 3.9 in [27]. First, we recall the following concentration theorem used by [27].

Theorem C.1 (Theorem 3.8 of [27]). *There is an absolute constant $C_{\mathrm{RV}} > 0$ such that the following holds. Let $Y_1, \ldots, Y_r$ be independent symmetric variables taking values in some Banach space. Assume $\|Y_j\| \leq R$ for all j, and let $Y := \|\sum_{i=1}^r Y_i\|$. Then, for any integers $l \geq Q$ and any $\tau > 0$, it holds that*

$$\Pr[Y \geq 8Q\mathbb{E}[Y] + 2Rl + \tau] \leq \Big(\frac{C_{\mathrm{RV}}}{Q}\Big)^l + 2\exp\Big(-\frac{\tau^2}{256Q\mathbb{E}[Y]^2}\Big).$$

From this theorem, we derive the following corollary.

Corollary C.2. *There are absolute constants $C_1, C_2 > 0$ such that the following holds. Let $Y_1, \ldots, Y_r$ be independent symmetric variables taking values in some Banach space. Assume $\|Y_j\| \leq R$ for all j, and let $Y := \|\sum_{i=1}^r Y_i\|$. Moreover, assume that $\mathbb{E}[Y] \leq E$ for some $E > 0$. Then, for every $\lambda \geq 1$, we have*

$$\Pr[Y \geq (C_1 + C_2\lambda)E] \leq 3\exp(-\lambda^2),$$

provided that $E \geq \lambda R$.

Proof. We properly set up the parameters of Theorem C.1. Let $\tau := 16\sqrt{Q}\lambda E$. Suppose $R > 0$ (otherwise, the conclusion is trivial). Let $Q := \lceil eC_{\mathrm{RV}} \rceil$ so that

$$\Big(\frac{C_{\mathrm{RV}}}{Q}\Big)^l \leq \exp(-l). \tag{C.1}$$

Let $l := Q\lceil \tau/(2R) \rceil = Q\lceil 8\sqrt{Q}\lambda E/R \rceil \geq \lambda^2$, where the inequality is because of the assumption $E/R \geq \lambda$. The coefficient Q also ensures that $l \geq Q$. Note that

$$R \leq E/\lambda \leq E\lambda \leq \tau \Rightarrow 2Rl \leq 2RQ(\tau/(2R) + 1) = Q\tau + 2QR \leq 3Q\tau. \tag{C.2}$$

Thus,

$$\Pr[Y \geq 8QE + 2Rl + \tau] \leq \Pr[Y \geq 8Q\mathbb{E}[Y] + 2Rl + \tau] \leq 3\exp(-\lambda^2),$$

where the second inequality follows from Theorem C.1 and by observing the choice of τ, the bound (C.1), and the lower bounds on l. Finally,

$$8QE + 2Rl + \tau \overset{(\mathrm{C.2})}{\leq} 8QE + (3Q+1)\tau = 8QE + 16(3Q+1)\sqrt{Q}\lambda E =: (C_1 + C_2\lambda)E,$$

where $C_1 := 8Q$ and $C_2 := 16(3Q+1)\sqrt{Q}$. The result now follows since

$$\Pr[Y \geq (C_1 + C_2\lambda)E] \leq \Pr[Y \geq 8QE + 2Rl + \tau]. \qquad \square$$

We now return to the proof of Lemma 4.5. In order to prove the desired tail bound, we shall apply Corollary C.2 on the norm of an independent summation of matrices. Recall that $N = q^{\tilde{k}}$. Let the variable $t \in \mathbb{F}_q^{\tilde{k}}$ be chosen uniformly at random, and consider the random $(q-1) \times N$ matrix $A := \varphi(\mathsf{Lin}_{\{t\}})$ formed by picking the tth row of the $N \times N$ matrix Lin and replacing each entry by a column vector representing its simplex encoding. Let $\mathcal{A} := A^\top A - (q-1)I_N$, where I_N is the $N \times N$ identity matrix, and let $\|\mathcal{A}\|_\Upsilon$ denote the following norm:

$$\|\mathcal{A}\|_\Upsilon := \sup_{x \in \mathcal{B}_2^{k,N}} |x^\top \mathcal{A} x|.$$

Denote the rows of A by $A_1, \ldots, A_{q-1}$, and observe that for every $x \in \mathcal{B}_2^{k,N}$ and $i \in \{1, \ldots, q-1\}$,

$$|\langle A_i, x\rangle| \leq \|A_i\|_\infty \|x\|_1 \leq \sqrt{k}, \tag{C.3}$$

where the second inequality follows from Cauchy–Schwarz. Therefore, since

$$\mathcal{A} = \sum_{i=1}^{q-1} (A_i^\top A_i - I_N)$$

for every $x \in \mathcal{B}_2^{k,N}$, we have

$$x^\top \mathcal{A} x = \sum_{i=1}^{q-1} \langle A_i, x\rangle^2 - (q-1) \overset{(C.3)}{\leq} (q-1)(k-1),$$

and thus

$$\|\mathcal{A}\|_\Upsilon \leq qk. \tag{C.4}$$

Suppose the original random row of Lin is written as a vector over $\mathbb{F}_q^N$ with coordinates indexed by the elements of $\mathbb{F}_q^{\tilde{k}}$. That is, $\mathsf{Lin}_{\{t\}} =: (w(u))_{u \in \mathbb{F}_q^{\tilde{k}}} =: w$. In particular, $w(u) = \langle u, t\rangle$, where the inner product is over $\mathbb{F}_q$. Let $u, v \in \mathbb{F}_q^{\tilde{k}}$. By basic linear algebra,

$$\Pr_t[w(u) = w(v)] = \Pr[\langle (u-v), t\rangle = 0] = \begin{cases} 1/q & \text{if } u \neq v, \\ 1 & \text{if } u = v. \end{cases}$$

Note that the (u, v)th entry of the matrix $A^\top A$ can be written as

$$(A^\top A)(u, v) = \langle \varphi(w(u)), \varphi(w(v))\rangle \overset{(2.1)}{=} \begin{cases} -1 & \text{if } w(u) \neq w(v), \\ q-1 & \text{if } w(u) = w(v). \end{cases}$$

Therefore, from this we can deduce that $\mathbb{E}[A^\top A] = (q-1)I_N$, or in other words, all entries of $\mathcal{A}$ are centered random variables, i.e., $\mathbb{E}[\mathcal{A}] = 0$.

Let $X_1, \ldots, X_{|T|}$ be independent random matrices, each distributed identically to $\mathcal{A}$, and consider the independent matrix summation

$$X := X_1 + \cdots + X_{|T|}.$$

Since each summand is a centered random variable, X is centered as well. Recall the random variables Δ_x and Δ from (4.3) and (4.4), and observe that Δ_x can be written as

$$\Delta_x = x^\top X x,$$

which in turn implies

$$\Delta = \|X\|_\Upsilon.$$

Thus, the assumption of the lemma implies that

$$\mathbb{E}[\|X\|_\Upsilon] \le \delta' |T|(q-1),$$

and proving a tail bound on Δ is equivalent to proving a tail bound on the norm of X. This can be done using Corollary C.2. However, the result cannot be directly applied to X since the X_i are centered but not symmetric for $q > 2$. As in [27], we use standard symmetrization techniques to overcome this issue. Namely, let $\mathcal{B}$ be the symmetrized version of $\mathcal{A}$ defined as

$$\mathcal{B} := \mathcal{A} - \mathcal{A}',$$

where $\mathcal{A}'$ is an independent matrix identically distributed to $\mathcal{A}$. Similar to X, define

$$Y := Y_1 + \cdots + Y_{|T|},$$

where the Y_i are independent and distributed identically to $\mathcal{B}$. As in the proof of Theorem 3.9 of [27], a simple application of Fubini and triangle inequalities implies that

$$\begin{gathered}\mathbb{E}[X] \le \mathbb{E}[Y] \le 2\mathbb{E}[X], \\ \Pr[X > 2\mathbb{E}[X] + \tau] \le 2\Pr[Y > \tau]. \end{gathered} \tag{C.5}$$

Let $E := 2\delta'|T|(q-1)$ so that by the above inequalities we know that $\mathbb{E}[Y] \le E$. We can now apply Corollary C.2 to Y and deduce that, for some absolute constants $C_1, C_2 > 0$, and every $\lambda \ge 1$,

$$\Pr[Y \ge (C_1 + C_2\lambda)E] \le \exp(-\lambda^2), \tag{C.6}$$

provided that $E \ge \lambda R$, where we can take $R = qk$ by (C.4). Plugging in the choice of E, we get the requirement that

$$\frac{|T|}{k} \ge \frac{\lambda q}{2\delta'(q-1)},$$

which can be ensured by an appropriate choice of c_3 in (4.36). Finally, (C.5) and (C.6) can be combined to deduce that

$$\begin{aligned}\Pr[X > 2E + (C_1 + C_2\lambda)E] &\le \Pr[X > 2\mathbb{E}[X] + (C_1 + C_2\lambda)E] \\ &\le 2\Pr[Y > (C_1 + C_2\lambda)E] \\ &\le 6\exp(-\lambda^2).\end{aligned}$$

This completes the proof of Lemma 4.5.

Acknowledgments. The authors would like to thank anonymous referees for their careful reviews and constructive comments, in particular for the suggestion of a direct proof of Theorem 3.5 that appears in Appendix B.

REFERENCES

[1] N. Ailon and E. Liberty, *An almost optimal unrestricted fast Johnson-Lindenstrauss transform*, in Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms, 2011.

[2] R. G. Baraniuk, M. A. Davenport, R. A. DeVore, and M. B. Wakin, *A simple proof of the restricted isometry property for random matrices*, Constr. Approx., 28 (2008), pp. 253–263.

[3] V. M. Blinovsky, *Bounds for codes in the case of list decoding of finite volume*, Probl. Inf. Transm., 22 (1986), pp. 7–19.

[4] V. M. Blinovsky, *On the convexity of one coding-theory function*, Probl. Inf. Transm., 44 (2008), pp. 34–39.

[5] E. Candès, *The restricted isometry property and its implications for compresses sensing*, C. R. Math. Acad. Sci. Paris, 346 (2008), pp. 589–592.

[6] E. Candès, J. Romberg, and T. Tao, *Robust uncertainty principle: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[7] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1208–1223.

[8] E. Candès and T. Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.

[9] B. Carl, *Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces*, Ann. Inst. Fourier, 35 (1985), pp. 79–118.

[10] M. Cheraghchi, *Applications of Derandomization Theory in Coding*, Ph.D. thesis, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, 2010; also available online from http://eccc.hpi-web.de/static/books/Applications_of_Derandomization_Theory_in_Coding/.

[11] M. Cheraghchi, *Coding-theoretic methods for sparse recovery*, in Proceedings of the Annual Allerton Conference on Communication, Control, and Computing, 2011.

[12] D. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[13] P. Elias, *Error-correcting codes for list decoding*, IEEE Trans. Inform. Theory, 37 (1991), pp. 5–12.

[14] O. Goldreich and L. Levin, *A hard-core predicate for all one-way functions*, in Proceedings of the 21st Annual ACM Symposium on Theory of Computing, 1989, pp. 25–32.

[15] P. Gopalan, V. Guruswami, and P. Raghavendra, *List decoding tensor products and interleaved codes*, SIAM J. Comput., 40 (2011), pp. 1432–1462.

[16] P. Gopalan, A. R. Klivans, and D. Zuckerman, *List-decoding Reed-Muller codes over small fields*, in Proceedings of the 40th Annual ACM Symposium on Theory of Computing, 2008, pp. 265–274.

[17] V. Guruswami, J. Håstad, and S. Kopparty, *On the list-decodability of random linear codes*, IEEE Trans. Inform. Theory, 57 (2011), pp. 718–725.

[18] V. Guruswami, J. Håstad, M. Sudan, and D. Zuckerman, *Combinatorial bounds for list decoding*, IEEE Trans. Inform. Theory, 48 (2002), pp. 1021–1035.

[19] V. Guruswami and S. Narayanan, *Combinatorial limitations of a strong form of list decoding*, Electron. Colloquium Computational Complexity, 19 (2012), p. 17.

[20] V. Guruswami and A. Smith, *Codes for computationally simple channels: Explicit constructions with optimal rate*, in Proceedings of the IEEE Symposium on the Foundations of Computer Science, 2010.

[21] V. Guruswami and M. Sudan, *Extensions to the Johnson Bound*, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.145.9405 (2001).

[22] V. Guruswami and S. Vadhan, *A lower bound on list size for list decoding*, IEEE Trans. Inform. Theory, 56 (2010), pp. 5681–5688.

[23] F. Krahmer and R. Ward, *New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property*, SIAM J. Math. Anal., 43 (2011), pp. 1269–1281.

[24] S. R. Kumar and D. Sivakumar, *Proofs, codes, and polynomial-time reducibilities*, in Proceedings of the 14th Annual IEEE Conference on Computation Complexity, 1999.

[25] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer-Verlag, Berlin, 1991.

[26] E. Mossel and C. Umans, *On the complexity of approximating the VC dimension*, J. Comput. System Sci., 65 (2001), pp. 660–671.

[27] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math., 61 (2008), pp. 1025–1045.

[28] A. Rudra, *Limits to list decoding of random codes*, IEEE Trans. Inform. Theory, 57 (2011), pp. 1398–1408.

[29] M. Sudan, L. Trevisan, and S. Vadhan, *Pseudorandom generators without the XOR lemma*, J. Comput. Systems Sci., 62 (2001), pp. 236–266.

[30] L. Trevisan, *Extractors and pseudorandom generators*, J. ACM, 48 (2001), pp. 860–879.

[31] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, in Compressed Sensing, Theory and Applications, Y. Eldar and G. Kutyniok, eds., Cambridge University Press, Cambridge, UK, 2012, pp. 210–268.

[32] M. Wootters, *On the list decodability of random linear codes with large error rate*, CoRR, abs/1302.2261, 2013.

[33] V. V. Zyablov and M. S. Pinsker, *List cascade decoding*, Probl. Inf. Transm., 17 (1981), pp. 236–240.