

Every list-decodable code for high noise has abundant near-optimal rate puncturings*

Atri Rudra[†] Mary Wootters[‡]

October 7, 2013

[†] University at Buffalo (SUNY)
atri@buffalo.edu

[‡] University of Michigan
wootters@umich.edu

Abstract

We show that any q -ary code with sufficiently good distance can be randomly punctured to obtain, with high probability, a code that is list decodable up to radius $1 - 1/q - \varepsilon$ with near-optimal rate and list sizes.

Our results imply that “most” Reed-Solomon codes are list decodable beyond the Johnson bound, settling the long-standing open question of whether *any* Reed Solomon codes meet this criterion. More precisely, we show that a Reed-Solomon code with random evaluation points is, with high probability, list decodable up to radius $1 - \varepsilon$ with list sizes $O(1/\varepsilon)$ and rate $\tilde{\Omega}(\varepsilon)$. As a second corollary of our argument, we obtain improved bounds on the list decodability of random linear codes over large fields.

Our approach exploits techniques from high dimensional probability. Previous work used similar tools to obtain bounds on the list decodability of random linear codes, but the bounds did not scale with the size of the alphabet. In this paper, we use a chaining argument to deal with large alphabet sizes.

*AR’s research is supported in part by NSF CAREER grant CCF-0844796 and NSF grant CCF-CCF-1161196. MW’s research is partially supported by the Simons Institute and a Rackham predoctoral fellowship. MW also thanks the Simons Institute for their hospitality.

1 Introduction

List decoding, proposed by Elias [Eli57] and Wozencraft [Woz58], is a relaxation of the traditional notion of unique decoding. In this relaxation, the decoder is allowed to output a small list of potentially transmitted messages with the guarantee that the transmitted codeword is in the list.

A remarkable fact about list decoding is that it effectively doubles the correctable fraction of errors. For any code over alphabet of size q , no more than a $\frac{1}{2} \left(1 - \frac{1}{q}\right)$ fraction of errors can be decoded uniquely. However, when the decoder may output a short list, there are codes which can tolerate a $1 - \frac{1}{q} - \varepsilon$ fraction of errors, for any $\varepsilon > 0$. This fact has been crucially exploited in numerous applications of list decoding in theoretical computer science and in particular, in complexity theory.¹ There are two important features of these applications:

1. Even though in the traditional communication setting it makes sense to consider constant fraction ρ of errors (in particular, ρ is close to 0), for complexity applications it is necessary for the fraction of correctable errors to be arbitrarily close to $1 - \frac{1}{q}$.
2. The optimal rate to correct $1 - \frac{1}{q} - \varepsilon$ fraction of errors is known, and is given by

$$R^*(q, \varepsilon) := 1 - H_q(1 - 1/q - \varepsilon) = \min \left\{ \varepsilon, \frac{q\varepsilon^2}{2 \log(q)} + O_q(\varepsilon^3) \right\}.$$

However, for complexity applications it is often enough to design a code with rate $\Omega(R^*(q, \varepsilon))$ with the same error correction capability.²

In this paper, we consider the list decoding problem in these parameter regimes. That is, we seek to correct a $1 - 1/q - \varepsilon$ fraction of errors, with rate $\Omega(R^*(q, \varepsilon))$ which may be suboptimal by multiplicative factors. The quest for such codes comes in two flavors: one can ask about the list decodability of a specific family of codes, or one can ask for the most general conditions which guarantee list decodability. This work addresses open problems of both flavors, discussed more below.

Specific families of codes with near-optimal rate. Many complexity applications require efficient correction of $1 - 1/q - \varepsilon$ fraction of errors, sometimes even with a local decoding algorithm. Thus, there has been significant effort directed at designing efficiently-decodable codes with optimal rate. The first non-trivial progress towards this goal was due to work of Sudan [Sud97] and Guruswami-Sudan [GS99] who showed that *Reed-Solomon* (RS) codes³ can be list decoded efficiently from $1 - \varepsilon$ fraction of errors with rate ε^2 . This matches the so-called *Johnson bound*, which relates the fraction of errors any code can combinatorially list decode (with small list size) to the distance of the code.

The work of Guruswami and Sudan held the record for seven years, during which RS codes enjoyed the best known tradeoff between rate and fraction of correctable errors. However, Parvaresh and Vardy showed that a variant of Reed-Solomon codes can beat the Johnson bound [PV05]. This was then improved by Guruswami and Rudra who achieved the optimal rate of ε with Folded Reed-Solomon codes [GR08]. Since then this optimal rate result has been achieved with other codes: derivative codes [GW13], multiplicity codes [Kop12], folded Algebraic Geometric (AG) codes [GX12] as well as *subcodes* of RS and AG codes [GX13]. There has also been a lot of recent work on reducing the runtime and list size for folded RS codes [GW13, DL12, GK13].

Even though many of the recent developments on list decoding are based on Reed-Solomon codes, there has been no non-trivial progress on the list decodability of Reed-Solomon codes themselves since the work of Guruswami-Sudan. This is true even if we only ask for combinatorial (not necessarily efficient) decoding

¹See the survey by Sudan [Sud00] and Guruswami's thesis [Gur04] for more on these applications.

²In fact in some applications even polynomial dependence on $R^*(q, \varepsilon)$ is sufficient.

³An RS code encodes a low-degree univariate polynomial f over \mathbb{F}_q as a list of evaluations $(f(\alpha_1), \dots, f(\alpha_n))$ for a predetermined set of $n \leq q$ *evaluation points* in \mathbb{F}_q .

guarantees, and even for rates only slightly beyond the Johnson bound. The question of whether or not Reed-Solomon codes can be list decoded beyond the Johnson bound was our main motivation for this work:

Question 1. *Are there Reed-Solomon codes which can be combinatorially list decoded from a $1 - \varepsilon$ fraction of errors, with rate $\omega(\varepsilon^2)$?*

This question, which has been well-studied, is interesting for several reasons. First, Reed-Solomon codes themselves are arguably the most well-studied codes in the literature. Secondly, there are complexity applications where one needs to be able to list decode Reed-Solomon codes in particular: e.g. the average-case hardness of the permanent [CPS99]. Finally, the Johnson bound is a natural barrier and it is interesting to ask whether it can be overcome by natural codes.⁴ It is known that Reed-Muller codes (which are generalizations of RS codes) can be list decoded beyond the Johnson bound [Gop10, GKZ08].

There have been some indications that Reed-Solomon codes might *not* be list decodable beyond the Johnson bound. Guruswami and Rudra [GR06] showed that for a generalization of list decoding called list recovery, the Johnson bound indeed gives the correct answer for RS codes. Further, Ben-Sasson et al. [BSKR10] showed that for RS code where the evaluation set is all of \mathbb{F}_q , the correct answer is close to the Johnson bound. In particular, they show that to correct $1 - \varepsilon$ fraction of errors with polynomial list sizes, the RS code with \mathbb{F}_q as its evaluation points cannot have rate better than $\varepsilon^{2-\gamma}$ for any constant $\gamma > 0$. However, this result leaves open the possibility that one could choose the evaluation points carefully and obtain an RS code which can be combinatorially list decoded significantly beyond the Johnson bound.

Resolving the above possibility has been open since [GS98]: see e.g. [Gur04, Rud07, Vad12] for explicit formulations of this question.

Large families of codes with near-optimal rate. While the work on list decodability of specific families of codes have typically also been accompanied with list decoding algorithms, combinatorial results have tended to focus on larger classes of codes. Two classic results along these lines are (i) that random (linear) codes have optimal rate with high probability, and (ii) the fact, following from the Johnson bound, that any code with distance $1 - 1/q - \varepsilon^2$ can be list decoded from $1 - 1/q - \varepsilon$ fraction of errors.

Results of the second type are attractive since they guarantee list decodability for any code, deterministically, as long as the code has large enough distance. Unfortunately, it is known that the Johnson bound is tight for some codes [GS03], and so we cannot obtain a stronger form of (ii). However, one can hope for a result of the first type for list decodability, based on distance. More specifically, it is plausible that most *puncturings* of a code with good distance can beat the Johnson bound.

Recently, Wootters [Woo13] obtained such a result for constant q . In particular, that work shows that any code with distance $1 - 1/q - \varepsilon^2$ has many puncturings of rate $\Omega(\varepsilon^2/\log q)$ that are list decodable from a $1 - 1/q - \varepsilon$ fraction of errors. This rate is optimal up to constant factors when q is small, but is far from the optimal bound of $R^*(q, \varepsilon)$ for larger values of q , even when q depends only on ε and is otherwise constant. This leads to our second motivating question, left open from [Woo13]:

Question 2. *Is it true that any code with distance $1 - 1/q - \varepsilon^2$ has many puncturings of rate $\tilde{\Omega}(R^*(q, \varepsilon))$ that can list decode from $1 - 1/q - \varepsilon$ fraction of errors?*

Our Results. In this work, we answer Questions 1 and 2 in the affirmative. Our main result addresses Question 2. We show that random puncturings of any code with distance $1 - 1/q - \varepsilon^2$ can list decode from $1 - 1/q - \varepsilon$ fraction of errors with rate

$$\frac{\min\{\varepsilon, q\varepsilon^2\}}{\log(q)\log^5(1/\varepsilon)}.$$

This improves upon the best known result in this regime by Wootters [Woo13] for $q \gtrsim \log^5(1/\varepsilon)$, and is optimal up to polylogarithmic factors. A corollary of this is that random linear codes are list decodable from

⁴We note that it is easy to come up with codes that have artificially small distance and hence can beat the Johnson bound.

$1 - 1/q - \varepsilon$ fraction of errors with the same rate—this improves the corresponding result in [Woo13] for the same range of parameters.

Our main result also implies a positive answer to Question 1, and we show that there do exist RS codes that are list decodable beyond the Johnson bound. In fact, most sets of evaluation points will work: we show that if an appropriate number of evaluation points are chosen at random, then with constant probability the resulting RS code is list decodable from $1 - \varepsilon$ fraction of errors with rate

$$\frac{\varepsilon}{\log(q) \log^5(1/\varepsilon)}.$$

This beats the Johnson bound for

$$\varepsilon \leq \tilde{O}\left(\frac{1}{\log(q)}\right).$$

Relationship to impossibility results. Before we get into the details, we digress a bit to explain why our result on Reed-Solomon codes does not contradict the known impossibility results on this question. The lower bound of [GR06] works for list recovery but does not apply to our results about list decoding.⁵ The lower bound of [BSKR10] does work for list decoding, but critically needs the set of evaluation points to be all of \mathbb{F}_q (or more precisely the evaluation set should contain particularly structured subsets \mathbb{F}_q). Since we pick the evaluation points at random, this property is no longer satisfied. Finally, Cheng and Wan [CW07] showed that *efficiently* solving the list decoding problem for RS codes from $1 - \varepsilon$ fraction of errors with rate $\Omega(\varepsilon)$ would imply an efficient algorithm to solve the discrete log problem. However, this result does not rule out the list size being small (which is what our results imply), just that algorithmically computing the list quickly is unlikely.

1.1 Approach and Organization

Our main technical result addresses Question 2 and states that a randomly punctured code⁶ will retain the list decoding properties of the original code as long as the original code has good distance. Our results for RS codes (answering Question 1) and random linear codes follow by starting from the RS code evaluated on all of \mathbb{F}_q and the q -ary Hadamard code, respectively.

After a brief overview of terminology in Section 2, we give a more detailed technical overview of our approach in Section 3. In Section 4 we state our main result, Theorem 2, about randomly punctured codes, and we apply it to Reed-Solomon codes and random linear codes. The remainder of the paper, Sections 5 and 6, are devoted to the proof of Theorem 2. Finally, we conclude with Section 7.

2 Preliminaries

Motivated by Reed-Solomon codes, we consider random ensembles of linear codes over \mathbb{F}_q , where the field size q is large. A code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is **linear** if it forms a subspace of \mathbb{F}_q^n . Equivalently, $\mathcal{C} = \{x^T G : x \in \mathbb{F}_q^k\}$ for a **generator matrix** $G \in \mathbb{F}_q^{k \times n}$. We refer to $x \in \mathbb{F}_q^k$ as the **message** and k as the **message length**. The length n of the resulting codeword $x^T G$ is called the **block length**.

We will study the list decodability of these codes, up to “large” error rates $1 - 1/q - \varepsilon$, which is $1 - \Theta(\varepsilon)$ when $q \gtrsim 1/\varepsilon$. We say that a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is (ρ, L) -**list decodable** if for all $z \in \mathbb{F}_q^n$, the number of codewords $c \in \mathcal{C}$ with $d(z, c) \leq \rho$ is at most L , where d denotes relative Hamming distance. We will actually study a slightly stronger notion of list decodability, explicitly studied in [GN13]. We say that a code $\mathcal{C} \subseteq \mathbb{F}_q^n$ is (ρ, L) -**average-radius list decodable** if for all $z \in \mathbb{F}_q^n$ and all sets Λ of $L + 1$ codewords $c \in \mathcal{C}$, the average

⁵Our results can be extended to the list recovery setting, and the resulting parameters obey the lower bound of [GR06].

⁶Technically, our construction is slightly different than randomly punctured codes: see Remark 3.

distance between elements of Λ and z is at least ρ . Notice that standard list decoding can be written in this language with the average replaced by a maximum.

In general, one is interested in the trade-off between ε , L , and the rate of the code \mathcal{C} . The **rate** of a linear code \mathcal{C} is defined to be $\dim(\mathcal{C})/n$, where $\dim(\mathcal{C})$ refers to the dimension of \mathcal{C} as a subspace of \mathbb{F}_q^n .

We'll consider ensembles of linear codes where the generator vectors are independent; this includes random linear codes and Reed Solomon codes with random evaluation points. More precisely, a distribution on the matrices G induces a distribution on linear codes. We say that such a distribution on linear codes \mathcal{C} has **independent symbols** if the columns of the generator matrix G are selected independently.

We will be especially interested in codes with randomly sampled symbols, where a new code (with a shorter block length) is created from an old code by including a few symbols of the codeword at random. Formally, suppose that \mathcal{C}' is a linear code over \mathbb{F}_q with generator matrix $G' \in \mathbb{F}_q^{k \times n'}$. Form a new generator matrix $G \in \mathbb{F}_q^{k \times n}$ whose columns are n columns of G' chosen independently at random (possibly with replacement). We say that the resulting random linear code \mathcal{C} with generator matrix G is a **randomly sampled** version of \mathcal{C}' , with block length n . Notice that randomly sampled codes have independent symbols by definition.

Remark 3 (Sampling vs. Puncturing). *We note that the operation of randomly sampling a code (a term we just made up) is very similar to that of randomly puncturing a code (a term with a long and illustrious history). The only difference is that we sample with replacement, while a randomly punctured code can be viewed as a code where the sampling is done without replacement. Our method of sampling is convenient for our analysis because of the independence. However, for the parameter regimes we will work in, collisions are overwhelmingly unlikely, and the distribution on randomly sampled codes is indeed very similar to that of randomly punctured codes.*

2.1 Notation

Throughout, we will consider linear codes $\mathcal{C} \subseteq \mathbb{F}_q^n$ of block length n and message length k , with generator matrices $G \in \mathbb{F}_q^{k \times n}$. The size of \mathcal{C} will be $|\mathcal{C}| = N$. For a message $x \in \mathbb{F}_q^k$, we will write $c = c(x)$ for the encoding $c(x) = x^T G$. We will be interested in subsets $\Lambda \subseteq \mathbb{F}_q^k$ of size L (the *list size*), which we will identify, when convenient, with the corresponding subset of \mathcal{C} .

For $x, y \in \mathbb{F}_q^n$, let $\text{agr}(x, y) = n(1 - d(x, y))$ be the number of symbols in which x and y agree. We will use $f(x) \lesssim g(x)$ (or $f(x) \gtrsim g(x)$) to indicate that there is some constant C so that $f(x) \leq Cg(x)$ (resp. $g(x) \leq Cf(x)$) for all x . Throughout, C_0, C_1, \dots and c_0, c_1, \dots will denote numerical constants. For clarity, we have made no attempt to optimize the constants. For a vector $v = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ and a set $S \subseteq [n]$, we will use v_S to denote the restriction of v to the coordinates indexed by S . We will use the ℓ_p norm $\|v\|_p = (\sum_{i=1}^n v_i^p)^{1/p}$, and the ℓ_∞ norm $\|v\|_\infty = \max_{j \in [n]} |v_j|$. We use \log to denote the logarithm base 2, and \ln to denote the natural log.

We will also use some machinery about Gaussian processes, but we have made an effort to keep this self-contained. For the reader's convenience, a few useful facts about Gaussian random variables are recorded in Appendix B. Finally, we will also use the following form of Chernoff(-Hoeffding) bound:

Theorem 1. *Let X_1, \dots, X_m be m independent random variables such that for every $i \in [m]$, $X_i \in [a_i, b_i]$, then for the random variable*

$$S = \sum_{i=1}^m X_i,$$

and any positive $v \geq 0$, we have

$$\mathbb{P}\{|S - \mathbb{E}[S]| \geq v\} \leq 2 \exp\left(-\frac{2v^2}{\sum_{i=1}^m (b_i - a_i)^2}\right).$$

3 Technical overview

In this section, we give a technical overview of our argument, and point out where it differs from previous approaches. The most similar argument in the literature is in [Woo13], which applies to random linear codes (but *not* Reed-Solomon codes). Below, we point out how our approach deviates, and where our improvements come from.

We first recall the classic proof of list decodability of general random codes. For a general random code, a Chernoff bound establishes that for a given Λ and z , there is only a very small probability that the codewords corresponding to Λ are all close to z . This probability is small enough to allow for a union bound over the $q^n \cdot \binom{N}{L}$ choices for Λ and z . However, this argument crucially exploits the independence between the encodings of distinct messages. If we begin with a random linear code (or a Reed-Solomon code with random evaluation points), then codewords are no longer independent, and the above argument fails. The classic way around this is to consider only the linearly independent messages in Λ ; however, this results in exponentially large list sizes of $q^{\Omega(1/\varepsilon)}$. The exponential dependence on ε can be removed for a *constant* fraction of errors, by a careful analysis of the dependence between codewords corresponding to linearly dependent messages [GHK11]. However, such techniques do not seem to work in the large-error regime that we consider.

In contrast, the approaches of [CGV13, Woo13] avoid analyzing the dependence between codewords by using tools from high dimensional probability. These arguments, which imply list decodability results for random linear codes, work when the error rate approaches $1 - 1/q$, and they (implicitly) use an improved union bound to avoid having to union bound over all Λ and z . However, these arguments do not scale well with q , which is crucial for the application to Reed-Solomon codes. In this work, we follow the approach of [Woo13] and use techniques from high dimensional probability and Gaussian processes to avoid the naive union bound. However, our arguments *will* scale with q , and thus are applicable to Reed-Solomon codes.

Following the approach of [Woo13], our proof actually establishes *average-radius* list decodability. The standard definition of list decodability has to do with bounding the maximum distance of a set $\Lambda \subseteq \mathcal{C}$ of L codewords from its centroid $z \in \mathbb{F}_q^n$. In contrast, average-radius list decodability is a stronger notion which focuses on the average distance from Λ to z .

The advantage of considering average-radius list decoding is that it linearizes the problem; after some rearranging (which is encapsulated in Proposition 1), it becomes sufficient to control

$$\sum_{c \in \Lambda} \text{agr}(z, c) = \sum_{c \in \Lambda} \sum_{j=1}^n \mathbf{1}_{c_j = z_j}$$

uniformly over all $\Lambda \subseteq \mathcal{C}$ and all $z \in \mathbb{F}_q^n$. We will show that this is true in expectation; that is, we will bound

$$\mathbb{E} \max_{\Lambda, z} \sum_{c \in \Lambda} \sum_{j=1}^n \mathbf{1}_{c_j = z_j}. \tag{1}$$

The proof proceeds in two steps.

The first (more straightforward) step is to argue that if the expectation and the maximum over Λ were reversed in (1), then we would have the control we need. To that end, we introduce a parameter

$$\mathcal{E} = \max_{|\Lambda|=L} \mathbb{E} \max_{z \in \mathbb{F}_q^n} \sum_{c \in \Lambda} \sum_{j=1}^n \mathbf{1}_{c_j = z_j}.$$

It is not hard to see that the received word z which maximizes the agreement is the one which, for each j , agrees with the plurality of the c_j for $c \in \Lambda$. That is,

$$\max_{z \in \mathbb{F}_q^n} \sum_{c \in \Lambda} \sum_{j=1}^n \mathbf{1}_{c_j = z_j} = \sum_{j=1}^n \max_{\alpha \in \mathbb{F}_q} |\{c \in \Lambda : c_j = \alpha\}| =: \sum_{j=1}^n \text{plurality}_j(\Lambda).$$

Thus, to control \mathcal{E} , we must understand the expected pluralities. For our applications, this follows from standard Johnson-bound type arguments.

Of course, it is generally not okay to switch expectations and maxima; we must also argue that the quantity inside the maximum does not deviate too much from its mean in the worst case. This is the second and more complicated step of our argument. We must control the deviation

$$\sum_{j=1}^n (\text{plurality}_j(\Lambda) - \mathbb{E}\text{plurality}_j(\Lambda)) \quad (2)$$

uniformly over all Λ of size L . By the assumption of independent symbols (that is, independently chosen evaluation points for the Reed-Solomon code, or independent generator vectors for random linear codes), each summand in (2) is independent.

Sums of independent random variables tend to be reasonably concentrated, but, as pointed out above, because the codewords are not independent there is no reason that the pluralities themselves need to be particularly well-concentrated. Thus, we cannot handle a union bound over all $\Lambda \subseteq \mathcal{C}$ of size L . Instead, we use a *chaining argument* to deal with the union bound. The intuition is that if the set Λ is close to the set Λ' (say they overlap significantly), then we should not have to union bound over both of them as though they were unrelated.

Our main theorem, Theorem 2, bounds the deviation (2), and thus bounds (1) in terms of \mathcal{E} . We control \mathcal{E} in the Corollaries 1 and 2, and then explain the consequences for Reed-Solomon codes and random linear codes in Sections 4.2 and 4.3.

We prove Theorem 2 in Section 5. To carry out the intuition above, we first pass to the language of Gaussian processes. Through some standard tricks from high dimensional probability, it will suffice to instead bound the Gaussian process

$$X(\Lambda) = \sum_{j=1}^n g_j \text{plurality}_j(\Lambda). \quad (3)$$

uniformly over all Λ of size L , where the g_j are independent standard normal random variables.

So far, this approach is similar to that of [Woo13]. The difference is that Wootters first maps the problem to \mathbb{C} , using a technique from [CGV13], in a way that allows for a slick bound on the relevant Gaussian process. However, this approach loses information about the size of q . In particular, the expected size of the pluralities decreases as q increases, and the approach of [Woo13] does not take advantage of this. In our approach, we deal with the pluralities directly, without embedding into \mathbb{C} . This unfortunately gives up on the slickness (our argument is somewhat technical), but allows us to take advantage of large q . We outline our methods below.

Returning to the Gaussian process (3), we condition on \mathcal{C} , considering only the randomness over the Gaussians. We control this process in Theorem 3, the proof of which is contained in Section 6. The process (3) induces a metric on the space of sets Λ : Λ is close to Λ' if the vectors of their pluralities are close, in ℓ_2 distance. Indeed, if Λ is close to Λ' in this sense, then the corresponding increment $X(\Lambda) - X(\Lambda')$ is small with high probability. In this language, the previous intuition about “wasting” the union bound on close-together Λ and Λ' can be made precise—for example, Dudley’s theorem [LT91, Tal05] bounds the supremum of the process in terms of the size of ε -nets with respect to this distance.

Thus, our proof of Theorem 3 boils down to constructing nets on the space of Λ ’s. In fact, our nets are quite simple—smaller nets consist of all of the sets of size $L/2^t$, for $t = 1, \dots, \log(L)$. However, showing that the width of these nets is small is trickier. Our argument actually uses the structure of the chaining argument that is at the heart of the proof of Dudley’s theorem: instead of arguing that the width of the net is small, we argue that each successive net cannot have points that are too far from the previous net, and thus build the “chain” step-by-step. One can of course abstract out a distance argument and apply Dudley’s theorem as a black-box. However, at the point that we are explicitly constructing the chains, we feel that it is more intuitive to include the entire argument. To this end, (and to keep the paper self-contained), we unwrap Dudley’s theorem in Section 6.2.

We construct and control our nets in Lemma 1, which we prove in Section 6.3. Briefly, the idea is as follows. In order to show that a set Λ of size $L/2^t$ is “close” to some set Λ' of size $L/2^{t+1}$, we use the probabilistic method. We choose a set $\Lambda' \subseteq \Lambda$ at random, and argue that in expectation (after some appropriate normalization), the two are “close.” Thus, the desired Λ' exists. However, the expected distance of Λ to Λ' in fact depends on the quantity

$$Q_t = \max_{|\Lambda|=L/2^t} \sum_{j=1}^n \text{plurality}_j(\Lambda).$$

For $t = 0$, this is the quantity that we were trying to control in the first place in (1). Carrying this quantity through our argument, we are able to solve for it at the end and obtain our bound.

Controlling Q_t for $t > 0$ requires a bit of delicacy. In particular, as defined above $Q_{\log(L)}$ is deterministically equal to n , which it turns out is too large for our applications. To deal with this, we actually chain over not just the Λ , but also the set of the symbols $j \in [n]$ that we consider. In fact, if we did not do this trick, we would recover (with some extra logarithmic factors) the result of [Woo13] for random linear codes.

We remark that our argument has a similar flavor to some existing arguments in other domains, for example [Rud97, RV08], where a quantity analogous to Q_0 arises, and where analogous nets will work. Our approach is slightly different (in particular, our proof of distance is structurally quite different), although it is possible that one could re-frame our argument to mimic those.

4 Main theorem

In this section, we state our main technical result, Theorem 2. To begin, we first give a slightly stronger sufficient condition for list decodability, called average-radius list decodability (defined above in Section 2). Average-radius list decodability has been explicitly studied before in [GN13] and was used in [Woo13] to prove upper bounds on the list decodability of ensembles of linear codes for constant-sized q . All of our results will actually show average-radius list decodability, and the following proposition shows that this will imply the standard notion of list decodability.

Proposition 1. *Suppose that*

$$\max_{z \in \mathbb{F}_q^n} \max_{\Lambda \subset \mathbb{F}_q^k, |\Lambda|=L} \sum_{x \in \Lambda} \text{agr}(c(x), z) < nL \left(\varepsilon + \frac{1}{q} \right).$$

Then \mathcal{C} is $(1 - 1/q - \varepsilon, L - 1)$ -list decodable.

Proof. By definition, \mathcal{C} is $(1 - 1/q - \varepsilon, L - 1)$ -list decodable if for any $z \in \mathbb{F}_q^n$ and any set $\Lambda \subset \mathbb{F}_q^n$ of size L , there is at least one message $x \in \Lambda$ so that $\text{agr}(c(x), z)$ is at most $n(\varepsilon + 1/q)$, that is, if

$$\max_{z \in \mathbb{F}_q^n} \max_{|\Lambda|=L} \min_{x \in \Lambda} \text{agr}(c(x), z) < n \left(\varepsilon + \frac{1}{q} \right).$$

Since the average is always larger than the minimum, it suffices for

$$\max_{z \in \mathbb{F}_q^n} \max_{|\Lambda|=L} \sum_{x \in \Lambda} \text{agr}(c(x), z) < Ln \left(\varepsilon + \frac{1}{q} \right),$$

as claimed. □

Our main theorem gives conditions on ensembles of linear codes under which $\mathbb{E} \max_{z, \Lambda} \sum_{x \in \Lambda} \text{agr}(c(x), z)$ is bounded. Thus, it gives conditions under which Proposition 1 holds.

Theorem 2. Fix $\varepsilon > 0$. Let \mathcal{C} be a random linear code with independent symbols. Let

$$\mathcal{E} = \max_{\Lambda \subset \mathbb{F}_q^k, |\Lambda|=L} \mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^k} \left(\sum_{x \in \Lambda} \text{agr}(c(x), z) \right).$$

Then

$$\mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^n} \max_{\Lambda \subset \mathbb{F}_q^k, |\Lambda|=L} \sum_{x \in \Lambda} \text{agr}(c(x), z) \leq \mathcal{E} + Y + \sqrt{\mathcal{E}Y},$$

where

$$Y = C_0 L \log(N) \log^5(L)$$

for an absolute constant C_0 .

Together with Proposition 1, Theorem 2 implies results about the list decodability of random linear codes with independent symbols, which we present next.

Remark 4. We have chosen the statement of the theorem which gives the best bounds for Reed-Solomon codes, where $q \gg L$ is a reasonable parameter regime. An inspection of the proof shows that we may replace one $\log(L)$ factor with $\min\{\log(L), \log(q)\}$.

4.1 Consequences of Theorem 2: list decodability of Reed-Solomon codes and random linear codes

In this section, we derive some consequences of Theorem 2 for randomly sampled codes, in terms of the distance of the original code. Our motivating examples are Reed-Solomon codes with random evaluation points, and random linear codes, which both fit into this framework. Indeed, Reed-Solomon codes with random evaluation points are obtained by sampling symbols from the Reed-Solomon code with block length $n = q$, and a random linear code is a randomly sampled Hadamard code. We'll discuss the implications and optimality for the two motivating examples below in Sections 4.2 and 4.3 respectively.

Our corollaries are split into two cases: the first holds for all q , but only yields the correct list size when q is small. The second holds for $q \gtrsim 1/\varepsilon^2$, and gives an improved list size in this regime. As discussed below in Section 4.3, our results are nearly optimal in both regimes.

First, we prove a result for intended for use with small q .

Corollary 1 (Small q). Let \mathcal{C}' be a linear code over \mathbb{F}_q with distance $1 - \frac{1}{q} - \frac{\varepsilon^2}{2}$. Suppose that

$$n \geq \frac{C_0 \log(N) \log^5(L)}{\min\{\varepsilon, q\varepsilon^2\}},$$

and choose \mathcal{C} to be a randomly sampled version of \mathcal{C}' , of block length n . Then, with constant probability over the choice of \mathcal{C} , the code \mathcal{C} is $(1 - 1/q - \varepsilon', 2/\varepsilon^2)$ -list decodable, where $\varepsilon' = (2 + \sqrt{2})\varepsilon$.

Corollary 1 holds for all values of q , but the list size $L \gtrsim \varepsilon^{-2}$ is suboptimal when $q \gtrsim 1/\varepsilon$. To that end, we include the following corollary, which holds when $q \gtrsim 1/\varepsilon^2$ and attains the “correct” list size.⁷

Corollary 2 (Large q). Suppose that $q > 1/\varepsilon^2$, and that ε is sufficiently small. Let \mathcal{C}' be a linear code over \mathbb{F}_q with distance $1 - \varepsilon^2$. Let

$$n \geq \frac{2C_0 \log(N) \log^5(L)}{\varepsilon},$$

and choose \mathcal{C} to be a randomly sampled version of \mathcal{C}' , of block length n . Then, with constant probability over the choice of \mathcal{C} , the code \mathcal{C} is $(1 - \varepsilon', 1/\varepsilon)$ -list decodable, where $\varepsilon' = 5\varepsilon$.

⁷As discussed below, we do not know good lower bounds on list sizes for large q ; by “correct” we mean matching the performance of a general random code.

The proofs of Corollaries 1 and 2 amount to controlling the worst expectation \mathcal{E} . This control follows from standard Johnson bound-type statements, and the proofs are given in Appendix A. Below, we discuss the consequences (and optimality) of these corollaries for Reed-Solomon codes and random linear codes.

Remark 5 (Average-radius list decodability). *We remark that the proofs of both Corollaries 1 and 2 go through Proposition 1, and thus actually show average-radius list decodability, not just list decodability. In particular, the applications to both Reed-Solomon codes and random linear codes hold under this stronger notion as well.*

4.2 Most Reed-Solomon codes are list-decodable beyond the Johnson bound

Our results imply that a Reed-Solomon code with random evaluation points is, with high probability, list decodable beyond the Johnson bound.

We briefly recall the definition of Reed-Solomon codes, and set notation for our discussion. Fix $q \geq n$, and an integer k , and let $\{\alpha_1, \dots, \alpha_n\} \subseteq \mathbb{F}_q$ be a list of “evaluation points.” The corresponding **Reed-Solomon code** $\mathcal{C} \subset \mathbb{F}_q^n$ encodes a polynomial (message) $f \in \mathbb{F}_q[x]$ of degree at most $k - 1$ as

$$c(f) = (f(\alpha_1), f(\alpha_2), \dots, f(\alpha_n)) \in \mathbb{F}_q^n.$$

Note that there are q^k polynomials of degree at most $k - 1$, and thus $|\mathcal{C}| = q^k$.

For Reed-Solomon codes, we are often interested in the parameter regime when $q \geq n$ is quite large. In particular, below we will be especially interested in the regime when $q \gg 1/\varepsilon^2$, and so we will use Corollary 2 for this application. To apply Corollary 2, let \mathcal{C}' be the Reed-Solomon code of block length q (that is, every point in \mathbb{F}_q is evaluated), and choose the n evaluation points $(\alpha_i)_{i=1}^n$ for \mathcal{C} independently from \mathbb{F}_q . We will choose the block length n so that

$$n \lesssim \frac{\log(N) \log^5(1/\varepsilon)}{\varepsilon}.$$

It is well known that the generator matrix for \mathcal{C} will have full rank. In the favorable case, the rate of \mathcal{C} is at least

$$R \gtrsim \frac{\varepsilon}{\log(q) \log^5(1/\varepsilon)}. \quad (4)$$

Before we investigate the result of Corollary 2, let us pause to observe what the Johnson bound predicts for \mathcal{C} . The distance of \mathcal{C} is exactly $1 - (k - 1)/n$. Indeed, any two polynomials of degree $k - 1$ agree on at most $k - 1$ points, and this is attained by, say, the zero polynomial and any polynomial with k distinct roots in $\{\alpha_1, \dots, \alpha_n\}$. Thus, letting $\varepsilon = (k - 1)/n$, the Johnson bound predicts that \mathcal{C} has rate ε , distance $1 - \varepsilon$, and is list decodable up to $1 - O(\sqrt{\varepsilon})$, with polynomial list sizes.

Now, we compare this to the result of Corollary 2. The distance of \mathcal{C}' is $1 - (k - 1)/q$, so as long as $q \gtrsim k/\varepsilon^2$, we may apply Corollary 2. Then, Corollary 2 implies that the resulting Reed-Solomon code \mathcal{C} has rate

$$\Omega\left(\frac{\varepsilon}{\log(q) \log^5(1/\varepsilon)}\right),$$

distance $1 - \varepsilon$, and is list decodable up to radius $1 - 5\varepsilon$, with list sizes at most $1/\varepsilon$.

In particular, the tolerable error rate may be as large as $1 - O(\varepsilon)$, rather than $1 - O(\sqrt{\varepsilon})$, and the rate suffers only by logarithmic factors.

4.3 Near-optimal bounds for random linear codes over large alphabets

In addition to implying that most Reed-Solomon codes are list decodable beyond the Johnson bound, Corollaries 1 and 2 provide the best known bounds on random linear codes over large fields. This improves the recent work of one of the authors in [Woo13] for large q ; further, our new results are tight up to logarithmic factors.

Regime	Best known rate for random linear codes	Upper bound on rate	Best known list size for random linear codes	Lower bound on list size
$q = \log^5(1/\varepsilon)$	$\frac{\varepsilon^2}{\log(q)}$, [Woo13]			
$q = 1/\varepsilon$	$\frac{q\varepsilon^2}{\log(q)\log^5(1/\varepsilon)}$ this work Cor. 1	$\frac{q\varepsilon^2}{\log(q)}$	$\frac{1}{\varepsilon^2}$ [CGV13, Woo13], this work Cor. 1	$\frac{1}{q^5\varepsilon^2}$ [GV10]
$q = 1/\varepsilon^2$	$\frac{\varepsilon}{\log(q)\log^5(1/\varepsilon)}$	$1 - H_q\left(1 - \frac{1}{q} - \varepsilon\right)$	$\frac{1}{\varepsilon}$	
$q = 2^{\Omega(1/\varepsilon)}$	this work Cors. 1, 2	ε	this work Cor. 2	

Figure 1: The state of affairs for q -ary random linear codes. Above, the list decoding radius is $1 - 1/q - \varepsilon$, and we have suppressed constant factors.

Suppose that \mathcal{C}' is the Hadamard code over \mathbb{F}_q of dimension k ; that is, the generator matrix of $\mathcal{C}' \in \mathbb{F}_q^{k \times q^k}$ has all the elements of \mathbb{F}_q^k as its columns. The relative distance of \mathcal{C}' is $1 - 1/q$, and so we may apply the corollaries with any $\varepsilon > 0$ that we choose.

To this end, fix $\varepsilon > 0$, and let \mathcal{C} be a randomly sampled version of \mathcal{C}' , of block length

$$n = \frac{2C_0 \log(q^k) \log^5(1/\varepsilon)}{\varepsilon}.$$

It is not hard to see that the generator matrix of \mathcal{C} will have full rank with high probability, and so the rate of \mathcal{C} will be at least

$$R = k/n = \frac{\min\{\varepsilon, q\varepsilon^2\}}{2C_0 \log(q) \log^5(1/\varepsilon)}. \quad (5)$$

By Corollary 1, \mathcal{C} is list decodable up to error radius $1 - 1/q - O(\varepsilon)$, with list sizes at most $2/\varepsilon^2$. When $q \gtrsim 1/\varepsilon^2$, Corollary 2 applies, and we get the same result with an improved list size of $1/\varepsilon$.

We compare these results to known results on random linear codes in Figure 1. The best known results on the list decodability of random linear codes, from [Woo13], state that a random linear code of rate on the order of $\varepsilon^2/\log(q)$ is $(1 - 1/q - \varepsilon, O(1/\varepsilon^2))$ -list decodable. This is optimal (up to constant factors) for constant q , but it is suboptimal for large q . In particular, the bound on the rate is surpassed by our bound (5) when $q \gtrsim \log^5(1/\varepsilon)$.

When the error rate is $1 - 1/q - \varepsilon$, the optimal information rate for list decodable codes is given by the list decoding capacity theorem, which implies that we must have $R \leq 1 - H_q(1 - 1/q - \varepsilon)$. This expression behaves differently for different parameter regimes; in particular, when $q \leq 1/\varepsilon$ and ε is sufficiently small, we have

$$1 - H_q(1 - 1/q - \varepsilon) = \frac{q\varepsilon^2}{2 \log(q)(1 - 1/q)} + O(\varepsilon^3),$$

while when $q \geq 2^{\Omega(1/\varepsilon)}$, the optimal rate is linear in ε . For the first of these two regimes—and indeed whenever $q \leq 1/\text{poly}(\varepsilon)$ —our bound (5) is optimal up to polylogarithmic factors in $1/\varepsilon$. In the second regime, when q is exponentially large, our bound slips by an additional factor of $\log(q)$.

For the $q \leq 1/\varepsilon^2$ regime, our list size of $1/\varepsilon^2$ matches existing results, and when q is constant it matches the lower bounds of [GV10]. For $q \geq 1/\varepsilon^2$, our list size of $1/\varepsilon$ is the best known. There is a large gap between the lower bound of [GV10] and our upper bounds for large q . However, there is evidence that the most of discrepancy is due to the difficulty of obtaining lower bounds on list sizes. Indeed, a (general) random code of rate $1 - H_q(1 - 1/q - \varepsilon) - 1/L$ is list-decodable with list size L , implying that $L = O(1/\varepsilon)$ is the correct answer for $q \gtrsim 1/\varepsilon$. Thus, while our bound seems like it is probably weak for q super-constant but smaller than $1/\varepsilon^2$, it seems correct for $q \gtrsim 1/\varepsilon^2$.

5 Proof of Theorem 2: reduction to Gaussian processes

In this section, we prove Theorem 2. For the reader's convenience, we restate the theorem here.

Theorem (Theorem 2, restated). *Fix $\varepsilon > 0$. Let \mathcal{C} be a random linear code with independent symbols. Let*

$$\mathcal{E} = \max_{\Lambda \subseteq \mathbb{F}_q^k, |\Lambda|=L} \mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^k} \left(\sum_{x \in \Lambda} \text{agr}(c(x), z) \right).$$

Then

$$\mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^n} \max_{\Lambda \subseteq \mathbb{F}_q^k, |\Lambda|=L} \sum_{x \in \Lambda} \text{agr}(c(x), z) \leq \mathcal{E} + Y + \sqrt{\mathcal{E}Y},$$

where

$$Y = C_0 L \log(N) \log^5(L)$$

for an absolute constant C_0 .

To begin, we introduce some notation.

Notation 1. For a set $\Lambda \subseteq \mathbb{F}_q^k$, let \mathbf{pl}_j denote the (fractional) plurality of index $j \in [n]$:

$$\mathbf{pl}_j(\Lambda) = \frac{1}{|\Lambda|} \max_{\alpha \in \mathbb{F}_q} |\{x \in \Lambda : c(x)_j = \alpha\}|.$$

For a set $I \subseteq [n]$, let

$$\mathbf{pl}_I(\Lambda) \in [0, 1]^n$$

be the vector $(\mathbf{pl}_j(\Lambda))_{j=1}^n$ restricted to the coordinates in I , with the remaining coordinates set to zero.

Rephrasing the goal in terms of our new notation, the quantity we wish to bound is

$$\mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^n} \max_{|\Lambda|=L} \sum_{x \in \Lambda} \text{agr}(c(x), z) = L \cdot \mathbb{E}_{\mathcal{C}} \max_{|\Lambda|=L} \sum_{j \in [n]} \mathbf{pl}_j(\Lambda). \quad (6)$$

Moving the expectation inside the maximum recovers the quantity

$$\mathcal{E} = L \cdot \max_{|\Lambda|=L} \mathbb{E}_{\mathcal{C}} \sum_{j \in [n]} \mathbf{pl}_j(\Lambda),$$

which appears in the statement of Theorem 2. Since Theorem 2 outsources a bound on \mathcal{E} to the user (in our case, Corollaries 1 and 2), we seek to control the worst deviation

$$\begin{aligned} \mathcal{F} &:= L \cdot \mathbb{E}_{\mathcal{C}} \max_{|\Lambda|=L} \left| \sum_{j \in [n]} \mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}} \sum_{j \in [n]} \mathbf{pl}_j(\Lambda) \right| \\ &= L \cdot \mathbb{E}_{\mathcal{C}} \max_{|\Lambda|=L} \left| \sum_{j \in [n]} (\mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}} \mathbf{pl}_j(\Lambda)) \right|. \end{aligned} \quad (7)$$

Indeed, let

$$Q = Q(\mathcal{C}) = \max_{|\Lambda|=L} \sum_{j \in [n]} \mathbf{pl}_j(\Lambda),$$

so that $L \cdot \mathbb{E}_{\mathcal{C}} Q$ is the quantity in (6). Then,

$$\begin{aligned} \mathbb{E}_{\mathcal{C}} Q &= \mathbb{E}_{\mathcal{C}} \max_{|\Lambda|=L} \left(\sum_{j \in [n]} \mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}} \sum_{j \in I} \mathbf{pl}_j(\Lambda) + \mathbb{E}_{\mathcal{C}} \sum_{j \in I} \mathbf{pl}_j(\Lambda) \right) \\ &\leq \mathbb{E}_{\mathcal{C}} \max_{|\Lambda|=L} \left| \sum_{j \in [n]} \mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}} \sum_{j \in I} \mathbf{pl}_j(\Lambda) \right| + \max_{|\Lambda|=L} \mathbb{E}_{\mathcal{C}} \sum_{j \in [n]} \mathbf{pl}_j(\Lambda) \\ &= \frac{1}{L} (\mathcal{F} + \mathcal{E}), \end{aligned} \tag{8}$$

so getting a handle on \mathcal{F} would be enough. With that in mind, we return our attention to (7). By the assumption of independent symbols, the summands in (7) are independent. By a standard symmetrization argument followed by a comparison argument (made precise in Appendix B.3 as Lemmas 4 and 5, respectively), we may bound

$$\frac{1}{L} \mathcal{F} = \mathbb{E}_{\mathcal{C}} \max_{|\Lambda|=L} \left| \sum_{j \in [n]} (\mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}} \mathbf{pl}_j(\Lambda)) \right| \tag{9}$$

$$\leq \sqrt{2\pi} \mathbb{E}_{\mathcal{C}} \mathbb{E}_g \max_{|\Lambda|=L} \left| \sum_{j \in [n]} g_j \mathbf{pl}_j(\Lambda) \right| \tag{10}$$

Above, g_j are independent standard normal random variables.

Let

$$\mathcal{S}_0 = \{[n]\} \times \{\Lambda \subset \mathbb{F}_q^k : |\Lambda| = L\}, \tag{11}$$

so that we wish to control

$$\mathbb{E}_{\mathcal{C}} \mathbb{E}_g \max_{(I, \Lambda) \in \mathcal{S}_0} \left| \sum_{j \in I} g_j \mathbf{pl}_j(\Lambda) \right|.$$

At this stage, maximizing I over the one-element collection $\{[n]\}$ may seem like a silly use of notation, but we will use the flexibility as the argument progresses.

Condition on the choice of \mathcal{C} until further notice, and consider only the randomness over the Gaussian random vector $g = (g_1, \dots, g_n)$. In particular, this fixes $Q = Q(\mathcal{C})$. In order to take advantage of (9), we will study the Gaussian process

$$X(I, \Lambda) = \sum_{j \in I} g_j \mathbf{pl}_j(\Lambda) \tag{12}$$

indexed by $(I, \Lambda) \in \mathcal{S}_0$. The bulk of the proof of Theorem 2 is the following theorem, which controls the expected supremum of $X(I, \Lambda)$, in terms of Q .

Theorem 3. *Condition on the choice of \mathcal{C} . Then*

$$\mathbb{E}_g \max_{(I, \Lambda) \in \mathcal{S}_0} |X(I, \Lambda)| \leq C_3 \sqrt{Q \log(N) \log^5(L)}$$

for some constant C_3 .

We will prove Theorem 3 in Section 6. First, let us show how it implies Theorem 2. By (9), and applying Theorem 3, we have

$$\begin{aligned} \mathcal{F} &\leq \sqrt{2\pi} L \mathbb{E}_{\mathcal{C}} \mathbb{E}_g \max_{(I, \Lambda) \in \mathcal{S}_0} \left| \sum_{j \in I} g_j v_j(z, \Lambda) \right| \\ &\leq C_3 \sqrt{2\pi} L \mathbb{E}_{\mathcal{C}} \left[\sqrt{Q \log(N) \log^5(L)} \right] \\ &\leq C_3 \sqrt{2\pi} L \sqrt{\mathbb{E}_{\mathcal{C}} Q \log(N) \log^5(L)} \end{aligned}$$

Using the fact (8) that $\mathbb{E}_{\mathcal{C}} Q \leq \frac{1}{L} (\mathcal{E} + \mathcal{F})$,

$$\begin{aligned} \mathcal{F} &\leq C_3 \sqrt{2\pi} \sqrt{L (\mathcal{E} + \mathcal{F}) \log(N) \log^5(L)} \\ &=: \sqrt{Y (\mathcal{E} + \mathcal{F})}, \end{aligned}$$

where

$$Y := C_3^2 2\pi L \log(N) \log^5(L).$$

Solving for \mathcal{F} , this implies that

$$\mathcal{F} \leq \frac{Y + \sqrt{Y^2 + 4Y\mathcal{E}}}{2} \leq Y + \sqrt{Y\mathcal{E}}.$$

Then, from (8) and the definition of Q (recall that $L \cdot \mathbb{E}_{\mathcal{C}} Q$ is the quantity in (6)),

$$\begin{aligned} \mathbb{E}_{\mathcal{C}} \max_{I, \Lambda} \sum_{x \in \Lambda} \text{agr}(c(x), z) &= L \mathbb{E}_{\mathcal{C}} Q \\ &\leq \mathcal{E} + \mathcal{F} \\ &\leq \mathcal{E} + Y + \sqrt{Y\mathcal{E}}, \end{aligned}$$

as claimed. This proves Theorem 2.

6 Proof of Theorem 3: controlling a Gaussian process

In this section, we prove Theorem 3. Recall that the goal was to control the Gaussian process (12) given by

$$X(I, \Lambda) = \sum_{j \in I} g_j \mathbf{pl}_j(\Lambda).$$

Recall also that we are conditioning on the choice of \mathcal{C} . Because of this, for notational convenience, we will identify $\Lambda \subset \mathbb{F}_q^k$ with the corresponding set of codewords $\{c(x) : x \in \Lambda\} \subset \mathcal{C}$. That is, for this section, we will imagine that $\Lambda \subset \mathcal{C}$ is a set of codewords.

Notation 2. *When the code \mathcal{C} is fixed (in particular, for the entirety of Section 6), we will identify $\Lambda \subset \mathbb{F}_q^k$ with $\Lambda \subset \mathcal{C}$, given by*

$$\Lambda \leftarrow \{c(x) : x \in \Lambda\}.$$

To control the Gaussian process (12), we will use a so-called ‘‘chaining argument.’’ That is, we will define a series of nets, $\mathcal{S}_t \subset 2^{[n]} \times 2^{\mathcal{C}}$ and write, for any $(I_0, \Lambda_0) \in \mathcal{S}_0$,

$$|X(I_0, \Lambda_0)| \leq \left(\sum_{t=0}^{t_{\max}-1} |X(\pi_t(I_0, \Lambda_0)) - X(\pi_{t+1}(I_0, \Lambda_0))| \right) + |X(\pi_{t_{\max}}(I_0, \Lambda_0))|,$$

where $\pi_t(I_0, \Lambda_0) \in \mathcal{S}_t$ will shortly be determined, and $\pi_0(I_0, \Lambda_0) = (I_0, \Lambda_0)$. Then we will argue that each step in this “chain” (that is, each summand in the first term) is small with high probability, and union bound over all possible chains.

For Gaussian processes, such chaining arguments come in standard packages, for example Dudley’s integral inequality [LT91], or Talagrand’s generic chaining inequality [Tal05]. However, we choose to unpack the argument for two reasons. The first is that our choice of nets is informed by the structure of the chaining argument, and so we feel it is clearer to define the nets in the context of the complete argument. The second reason is to make the exposition self-contained.

We remark that, due to the nature of our argument, it is convenient for us to start with the large nets indexed by small t , and the small nets indexed by large t ; this is in contrast with convention.

6.1 Defining the nets

We will define nets \mathcal{S}_t , for each t recursively. Begin by defining \mathcal{S}_0 as in (11), and let $\pi_0 : \mathcal{S}_0 \rightarrow \mathcal{S}_0$ be the identity map. Given \mathcal{S}_t , we will define \mathcal{S}_{t+1} , as well as the maps $\pi_{t+1} : \mathcal{S}_0 \rightarrow \mathcal{S}_{t+1}$. Our maps π_t will satisfy the guarantees of the following lemma.

Lemma 1. *Fix a parameter $\eta = 1/\log(L)$, and suppose $c_0 < L < N/2$ is sufficiently large, for some constant c_0 . Let*

$$t_{\max} = \frac{\log(L) - 2\log(1/\eta) - 2}{\log(2/(1-\eta))}. \quad (13)$$

Then there is a sequence of maps

$$\pi_t : \mathcal{S}_0 \rightarrow 2^{\lfloor n \rfloor} \times 2^{\mathcal{C}}$$

for $t = 0, \dots, t_{\max}$ so that π_0 is the identity map and so that the following hold.

First, for all $(I_0, \Lambda_0) \in \mathcal{S}_0$, and for all $t = 0, \dots, t_{\max}$, the pair $(I_t, \Lambda_t) = \pi_t(I_0, \Lambda_0)$ obeys

$$\sum_{j \in I_t} \mathbf{pl}_j(\Lambda_t) \leq Q_t := (1 + \eta)^t Q. \quad (14)$$

and

$$\left(\frac{1-\eta}{2}\right)^t L \leq |\Lambda_t| \leq \left(\frac{1+\eta}{2}\right)^t L. \quad (15)$$

In addition, for all $(I_0, \Lambda_0) \in \mathcal{S}_0$, and for all $t = 0, \dots, t_{\max} - 1$, the pair $(I_{t+1}, \Lambda_{t+1}) = \pi_{t+1}(I_0, \Lambda_0)$ obeys

$$\left\| \mathbf{pl}_{I_t}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1}) \right\|_2 \leq \frac{C_4 \sqrt{Q_t \log(L)}}{\eta \sqrt{|\Lambda_t|}} \quad (16)$$

for some constant C_4 .

Finally, for all $t = 0, \dots, t_{\max}$, define

$$\mathcal{S}_t := \{\pi_t(I_0, \Lambda_0) : (I_0, \Lambda_0) \in \mathcal{S}_0\}.$$

Then, for $t \geq 1$, the size of the net \mathcal{S}_t satisfies

$$|\mathcal{S}_t| \leq C_6 \binom{N}{eL/2^t} \binom{N}{eL/2^{t-1}}, \quad (17)$$

for some constant C_6 , while $|\mathcal{S}_0| = \binom{N}{L}$.

6.2 Proof of Theorem 3 from Lemma 1: a chaining argument

Before we prove Lemma 1, we will show how to use it to prove Theorem 3. This part of the proof follows the standard proof of Dudley's theorem [LT91], and can be skipped by the reader already familiar with it.⁸ As outlined above, we will use a chaining argument to control the Gaussian process in Theorem 3. We wish to control

$$\mathbb{E} \max_{(I, \Lambda) \in \mathcal{S}_0} |X(I, \Lambda)|.$$

For any $(I_0, \Lambda_0) \in \mathcal{S}_0$, write

$$\begin{aligned} |X(I_0, \Lambda_0)| &\leq \left(\sum_{t=0}^{t_{\max}-1} |X(\pi_t(I_0, \Lambda_0)) - X(\pi_{t+1}(I_0, \Lambda_0))| \right) + |X(\pi_{t_{\max}}(I_0, \Lambda_0))| \\ &=: S(I_0, \Lambda_0) + |X(\pi_{t_{\max}}(I_0, \Lambda_0))|, \end{aligned} \quad (18)$$

where Lemma 1 tells us how to pick $(I_t, \Lambda_t) := \pi_t(I_0, \Lambda_0)$, and where we have used the fact that $\pi_0(I_0, \Lambda_0) = (I_0, \Lambda_0)$.

Each increment

$$X(\pi_t(I_0, \Lambda_0)) - X(\pi_{t+1}(I_0, \Lambda_0)) = \sum_{j=1}^n g_j [\mathbf{1}_{j \in I_t} \mathbf{pl}_j(\Lambda_t) - \mathbf{1}_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_{t+1})]$$

is a Gaussian random variable (see Fact 6 in Appendix B) with variance

$$\begin{aligned} \sum_{j=1}^n (\mathbf{1}_{j \in I_t} \mathbf{pl}_j(\Lambda_t) - \mathbf{1}_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_{t+1}))^2 &= \left\| \mathbf{pl}_{I_t}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1}) \right\|_2^2 \\ &\leq \frac{C_4^2 Q_t \log(L)}{\eta^2 |\Lambda_t|} \quad \text{by (16)} \\ &\leq \frac{C_4^2 Q_t \log(L)}{\eta^2 \left(\frac{1-\eta}{2}\right)^t L} \quad \text{by (15)} \\ &\leq \frac{C_4^2 (1+\eta)^t Q \log(L)}{\eta^2 \left(\frac{1-\eta}{2}\right)^t L} \quad \text{by (14)} \\ &\leq \left(\frac{C_4}{\eta}\right)^2 \left(\frac{Q \log(L) (2(1+2\eta))^t}{L}\right) \quad \text{using } \eta \leq 1/2. \\ &\leq \left(\frac{eC_4}{\eta}\right)^2 \left(\frac{Q \log(L) 2^t}{L}\right) \quad \text{using } \eta = 1/\log(L) \text{ and } t_{\max} \leq \log(L). \end{aligned}$$

Thus, for each $0 \leq t < t_{\max}$, and for any $u, a_t \geq 0$,

$$\begin{aligned} \mathbb{P} \{ |X(\pi_t(z, \Lambda)) - X(\pi_{t+1}(z, \Lambda))| > u \cdot a_t \} &\leq \exp \left(\frac{-u^2 \cdot a_t^2}{2 \sum_{j=1}^n (\mathbf{1}_{j \in I_t} \mathbf{pl}_j(\Lambda_t) - \mathbf{1}_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_{t+1}))^2} \right) \\ &\leq \exp \left(\frac{-u^2 \cdot a_t^2}{2 \left(\frac{eC_4}{\eta}\right)^2 \left(\frac{Q \log(L) 2^t}{L}\right)} \right) \\ &=: \exp \left(\frac{-u^2 \cdot a_t^2}{\delta_t^2} \right). \end{aligned} \quad (19)$$

⁸Assuming that the reader is willing to take our word on the calculations.

In the above, we used the fact that for a Gaussian variable g with variance σ , $\mathbb{P}\{|g| > u\} \leq \exp(-u^2/(2\sigma^2))$. Now we union bound over all possible “chains” (that is, sequences $\{\pi_t(I_0, \Lambda_0)\}_t$) to bound the probability that there exists a $(I_0, \Lambda_0) \in \mathcal{S}_0$ so that the first term $S(I_0, \Lambda_0)$ in (18) is large. Consider the event that for all $(I_0, \Lambda_0) \in \mathcal{S}_0$,

$$|X(\pi_t(I_0, \Lambda_0)) - X(\pi_{t+1}(I_0, \Lambda_0))| \leq u \cdot a_t,$$

for a_t to be determined shortly. In the favorable case that this event occurs, the first term in (18) is bounded by

$$S(I_0, \Lambda_0) = \sum_{t=0}^{t_{\max}-1} |X(\pi_t(I_0, \Lambda_0)) - X(\pi_{t+1}(I_0, \Lambda_0))| \leq u \cdot \sum_{t=0}^{t_{\max}-1} a_t,$$

for all (I_0, Λ_0) . Let

$$N_t = \begin{cases} C_6 \binom{N}{eL/2^t} \binom{N}{eL/2^{t-1}} & t \geq 1 \\ \binom{N}{L} & t = 0 \end{cases} \quad (20)$$

be our bound on $|\mathcal{S}_t|$, given by (17) in Lemma 1. Then probability that the above good event fails to occur is at most, by the union bound,

$$\mathbb{P} \left\{ \max_{(I_0, \Lambda_0) \in \mathcal{S}_0} S(I_0, \Lambda_0) > u \cdot \sum_{t=0}^{t_{\max}-1} a_t \right\} \leq \sum_{t=0}^{t_{\max}-1} N_t N_{t+1} \exp \left(\frac{-u^2 \cdot a_t^2}{\delta_t^2} \right).$$

Indeed, there are at most $N_t N_{t+1}$ possible “steps” between $\pi_t(I_0, \Lambda_0)$ and $\pi_{t+1}(I_0, \Lambda_0)$, and the probability that any step at level t fails is given by (19).

Choose

$$a_t = \sqrt{2 \ln(N_t N_{t+1})} \delta_t. \quad (21)$$

This choice will imply that

$$\mathbb{E} \max_{(I_0, \Lambda_0) \in \mathcal{S}_0} S(I_0, \Lambda_0) \leq 2 \sum_{t=1}^{t_{\max}-1} a_t. \quad (22)$$

For the reader’s convenience, a brief (standard) proof of (22) is included in Appendix B.2. Plugging in our definition (21) of a_t and then of δ_t and N_t (Equations (19) and (20), respectively),

$$\begin{aligned} \mathbb{E} \max_{(z, \Lambda) \in \mathcal{S}_0} S(I_0, \Lambda_0) &\leq 2 \sum_{t=0}^{t_{\max}-1} \sqrt{2 \ln(N_t N_{t+1})} \delta_t \\ &\lesssim \sum_{t=0}^{t_{\max}-1} \sqrt{\frac{L}{2^t} \log(N)} \left(\frac{1}{\eta} \sqrt{\frac{Q \log(L) 2^t}{L}} \right) \\ &= t_{\max} \left(\frac{\sqrt{Q \log(N) \log(L)}}{\eta} \right) \\ &\leq \log^2(L) \sqrt{Q \log(N) \log(L)}, \end{aligned} \quad (23)$$

after using the choice of $\eta = 1/\log(L)$ and $t_{\max} \leq \log(L)$ in the final line.

With the first term $S(I_0, \Lambda_0)$ of (18) under control by (23), we turn to the second term, and we now bound the probability that the final term $X(\pi_{t_{\max}}(z, \Lambda))$ is large. Let $(I_{\max}, \Lambda_{\max}) = \pi_{t_{\max}}(I_0, \Lambda_0)$, so we wish to bound the Gaussian random variable

$$X(\pi_{t_{\max}}(I_0, \Lambda_0)) = \sum_{j \in I_{\max}} g_j \mathbf{pl}_j(\Lambda_{\max}).$$

As with the increments in $S(I_0, \Lambda_0)$, we will first bound the variance of $X(\pi_{t_{\max}}(I_0, \Lambda_0))$. By (14), we know that

$$\sum_{j \in I_{\max}} \mathbf{pl}_j(\Lambda_{\max}) \leq Q_{t_{\max}} \leq eQ.$$

Further, since $\mathbf{pl}_j(\Lambda_{\max})$ is a fraction, we always have

$$\mathbf{pl}_j(\Lambda_{\max}) \leq 1.$$

By Hölder's inequality,

$$\sum_{j \in I_{\max}} \mathbf{pl}_j(\Lambda_{\max})^2 \leq \left(\sum_{j \in I_{\max}} \mathbf{pl}_j(\Lambda_{\max}) \right) \left(\max_{j \in I_{\max}} \mathbf{pl}_j(\Lambda_{\max}) \right) \leq eQ.$$

Thus, for each $(I_0, \Lambda_0) \in \mathcal{S}_0$, $X(\pi_{t_{\max}}(I_0, \Lambda_0))$ is a Gaussian random variable with variance at most eQ (using Fact 6 in Appendix B). We recall the choice from (13) of

$$t_{\max} = \frac{\log(L) - 2 \log(1/\eta) - 2}{1 + \log(1/(1 - \eta))} \geq \log(L) - 2 \log \log(L) - C_7, \quad (24)$$

for some constant C_7 , for sufficiently large L . Because there are $|\mathcal{S}_{t_{\max}}| \leq \binom{N}{eL/2^{t_{\max}}}$ of these, a standard estimate for the maximum of Gaussians (see Proposition 2 in Appendix B) gives

$$\begin{aligned} \mathbb{E} \max_{(I_0, \Lambda_0) \in \mathcal{S}_0} |X(\pi_{t_{\max}}(I_0, \Lambda_0))| &\lesssim \sqrt{\ln |\mathcal{S}_{t_{\max}}|} \cdot \sqrt{Q} \\ &\lesssim \sqrt{\frac{LQ \log(N)}{2^{t_{\max}}}} \\ &\lesssim \log(L) \sqrt{Q \log(N)}, \end{aligned}$$

using the choice of t_{\max} (and the bound on it in (24)) in the final line. Finally, putting together the two parts of (18), we have

$$\mathbb{E} \max_{(I_0, \Lambda_0) \in \mathcal{S}_0} X(I_0, \Lambda_0) \lesssim \log^2(L) \sqrt{Q \log(N) \log(L)} + \log(L) \sqrt{Q \log(N)} \lesssim \log^2(L) \sqrt{Q \log(N) \log(L)}. \quad (25)$$

This completes the proof of Theorem 3 (assuming Lemma 1).

6.3 Proof of Lemma 1: the desired nets exist

Finally, we prove Lemma 1. We proceed inductively. In addition to the conclusions of the lemma, we will maintain the inductive hypotheses

$$I_{t+1} \subseteq I_t \quad \text{and} \quad \Lambda_{t+1} \subseteq \Lambda_t \quad (26)$$

for all t .

For the base case, $t = 0$, we set $\pi_0(I_0, \Lambda_0) = (I_0, \Lambda_0)$. The definition of Q guarantees (14), and the definition of \mathcal{S}_0 guarantees (15). By definition $|\mathcal{S}_0| \leq \binom{N}{L}$. Further, since by definition $I_0 = [n]$, the first part of (26) is automatically satisfied. (We are not yet in a position to verify the base case for the second part of (26), having not yet defined Λ_1 , but we will do so shortly).

We will need to keep track of how the pluralities $\mathbf{pl}_j(\Lambda_t)$ change, and for this we need the following notation.

Notation 3. For $\alpha \in \mathbb{F}_q$ and $\Lambda \subset \mathcal{C}$, let

$$v_j(\alpha, \Lambda) = \frac{|\{c \in \Lambda : c_j = \alpha\}|}{|\Lambda|}$$

be the fraction of times the symbol α appears in the j 'th symbol in Λ .

Now we define \mathcal{S}_t for $t \geq 1$. Suppose we are given $(I_t, \Lambda_t) = \pi_t(I_0, \Lambda_0) \in \mathcal{S}_t$ satisfying the hypotheses of the lemma. We need to produce $(I_{t+1}, \Lambda_{t+1}) \in \mathcal{S}_{t+1}$, and we will use the probabilistic method. We will choose I_{t+1} deterministically based on Λ_t . Then we will choose Λ_{t+1} randomly, based on Λ_t , and show that with positive probability, (I_{t+1}, Λ_{t+1}) obey the desired conclusions. Then we will fix a favorable draw of (I_{t+1}, Λ_{t+1}) and call it $\pi_{t+1}(I_0, \Lambda_0)$.

We choose I_{t+1} to be the “heavy” coordinates,

$$I_{t+1} := \{j : |\Lambda_t| \mathbf{pl}_j(\Lambda_t) \geq \gamma\},$$

for

$$\gamma := \frac{4c_1 \log(L)}{(1-\eta)^2 \eta^2}, \quad (27)$$

where c_1 is a suitably large constant to be fixed later. Notice that I_{t+1} depends only on Λ_t (and on \mathcal{C} , which for the moment is fixed).

Now consider drawing $\Lambda_{t+1} \subset \Lambda_t$ at random by including each element of Λ_t in Λ_{t+1} independently with probability $1/2$. We will choose some Λ_{t+1} from the support of this distribution.

Before we fix Λ_{t+1} , observe that we are already in a position to establish (26). Indeed, the second part of (26) holds for all t , because $\Lambda_{t+1} \subseteq \Lambda_t$ by construction. To establish the first part of (26) for $t, t+1$, we use that $\Lambda_t \subseteq \Lambda_{t-1}$ (by induction, using (26) for $t-1, t$), and this implies that for all $j \in I_{t+1}$,

$$\begin{aligned} \gamma &\leq |\Lambda_t| \mathbf{pl}_j(\Lambda_t) \\ &= \max_{\alpha} |\{c \in \Lambda_t : c_j = \alpha\}| \\ &\leq \max_{\alpha} |\{c \in \Lambda_{t-1} : c_j = \alpha\}| \\ &= |\Lambda_{t-1}| \mathbf{pl}_j(\Lambda_{t-1}), \end{aligned}$$

and hence $j \in I_t$. Thus,

$$I_{t+1} \subseteq I_t. \quad (28)$$

Before we move on to the other inductive hypotheses, stated in Lemma 1, we must fix a “favorable” draw of Λ_{t+1} . In expectation, Λ_{t+1} behaves like Λ_t , and so the hope is that the “step”

$$\mathbf{pl}_{I_t}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1})$$

is small. We quantify this in the following lemma.

Lemma 2. *For all j ,*

$$\mathbb{E} [|\Lambda_{t+1}| |\mathbf{pl}_j(\Lambda_t) - \mathbf{pl}_j(\Lambda_{t+1})|] \leq \sqrt{C_5 |\Lambda_t| \log(L) \mathbf{pl}_j(\Lambda_t)}$$

and

$$\mathbb{E} [|\Lambda_{t+1}|^2 (\mathbf{pl}_j(\Lambda_t) - \mathbf{pl}_j(\Lambda_{t+1}))^2] \leq C_5 |\Lambda_t| \log(L) \mathbf{pl}_j(\Lambda_t)$$

for some constant C_5 .

Proof. The second statement implies the first, by Jensen’s inequality, so we prove only the second statement.

For each $\alpha \in \mathbb{F}_q$, and each $j \in [n]$, consider the random variable

$$\begin{aligned}
Y_j(\alpha) &:= |\Lambda_{t+1}| (v_j(\alpha, \Lambda_{t+1}) - v_j(\alpha, \Lambda_t)) \\
&= \sum_{c \in \Lambda_t: c_j = \alpha} \left(\xi_c - \frac{|\Lambda_{t+1}|}{|\Lambda_t|} \right) \\
&= \sum_{c \in \Lambda_t: c_j = \alpha} \left(\xi_c - \frac{1}{2} \right) + \sum_{c \in \Lambda_t: c_j = \alpha} \left(\frac{1}{2} - \frac{|\Lambda_{t+1}|}{|\Lambda_t|} \right) \\
&= \sum_{c \in \Lambda_t: c_j = \alpha} \left(\xi_c - \frac{1}{2} \right) + v_j(\alpha, \Lambda_t) \sum_{c \in \Lambda_t} \left(\frac{1}{2} - \xi_c \right) \\
&=: Z_j(\alpha) + W_j(\alpha),
\end{aligned}$$

where above ξ_c is 1 if $c \in \Lambda_{t+1}$ and 0 otherwise. Both $Z_j(\alpha)$ and $W_j(\alpha)$ are sums of independent mean-zero random variables, and we use Chernoff bounds to control them. First, $Z_j(\alpha)$ is a sum of $|\Lambda_t|v_j(\alpha, \Lambda_t)$ independent mean-zero random variables, and a Chernoff bound (Theorem 1) yields

$$\mathbb{P}\{|Z_j(\alpha)| > u\} \leq 2 \exp\left(\frac{-2u^2}{|\Lambda_t|v_j(\alpha, \Lambda_t)}\right) \leq 2 \exp\left(\frac{-2u^2}{|\Lambda_t|\mathbf{pl}_j(\Lambda_t)}\right).$$

Similarly, $W_j(\alpha)$ is a sum of $|\Lambda_t|$ independent mean-zero random variables, each contained in

$$\left[-\frac{v_j(\alpha, \Lambda_t)}{2}, \frac{v_j(\alpha, \Lambda_t)}{2}\right] \subseteq \left[-\frac{\mathbf{pl}_j(\Lambda_t)}{2}, \frac{\mathbf{pl}_j(\Lambda_t)}{2}\right],$$

and we have

$$\mathbb{P}\{|W_j(\alpha)| > u\} \leq 2 \exp\left(\frac{-2u^2}{|\Lambda_t|\mathbf{pl}_j(\Lambda_t)^2}\right) \leq 2 \exp\left(\frac{-2u^2}{|\Lambda_t|\mathbf{pl}_j(\Lambda_t)}\right),$$

using the fact that $\mathbf{pl}_j(\Lambda_t) \leq 1$. Together,

$$\mathbb{P}\{|Y_j(\alpha)| > u\} \leq \mathbb{P}\{|W_j(\alpha)| > u/2\} + \mathbb{P}\{|Z_j(\alpha)| > u/2\} \leq 4 \exp\left(\frac{-u^2}{2\mathbf{pl}_j(\Lambda_t)|\Lambda_t|}\right),$$

Let

$$T_j = \{\alpha \in \mathbb{F}_q : \exists c \in \Lambda_t, c_j = \alpha\}$$

be the set of symbols that show up in the j 'th coordinates of Λ_t . Then

$$|T_j| \leq \min\{q, |\Lambda_t|\} \leq L.$$

By the union bound, and letting $v = u^2$,

$$\mathbb{P}\left\{\max_{\alpha \in \mathbb{F}_q} Y_j(\alpha)^2 > v\right\} = \mathbb{P}\left\{\max_{\alpha \in T_j} Y_j(\alpha)^2 > v\right\} \leq 4L \exp\left(\frac{-v}{2\mathbf{pl}_j(\Lambda_t)|\Lambda_t|}\right). \quad (29)$$

Next, we show that if all of the $Y_j(\alpha)$ are under control, then so are the pluralities $\mathbf{pl}_j(\Lambda_t)$. For any four numbers A, B, C, D with $A \leq B$ and $C \leq D$, we have

$$|B - D| \leq \max\{|B - C|, |D - A|\}. \quad (30)$$

Indeed, we have

$$B - D \leq (B - D) + (D - C) = B - C \quad \text{and} \quad D - B \leq (D - B) + (B - A) = D - A.$$

The claim (30) follows. Now, for fixed j , let

$$\alpha = \operatorname{argmax}_{\sigma \in T_j} v_j(\sigma, \Lambda_t) \quad \text{and} \quad \beta = \operatorname{argmax}_{\sigma \in T_j} v_j(\sigma, \Lambda_{t+1}),$$

so that

$$|\Lambda_{t+1}|v_j(\alpha, \Lambda_{t+1}) \leq |\Lambda_{t+1}|v_j(\beta, \Lambda_{t+1}) \quad \text{and} \quad |\Lambda_{t+1}|v_j(\beta, \Lambda_t) \leq |\Lambda_{t+1}|v_j(\alpha, \Lambda_t).$$

By (30), we have

$$\begin{aligned} |\Lambda_{t+1}| |\mathbf{pl}_j(\Lambda_{t+1}) - \mathbf{pl}_j(\Lambda_t)| &= |\Lambda_{t+1}| |v_j(\beta, \Lambda_{t+1}) - v_j(\alpha, \Lambda_t)| \\ &\leq |\Lambda_{t+1}| \max \{ |v_j(\alpha, \Lambda_t) - v_j(\alpha, \Lambda_{t+1})|, |v_j(\beta, \Lambda_t) - v_j(\beta, \Lambda_{t+1})| \} \\ &\leq \max_{\alpha \in T_j} |Y_j(\alpha)|. \end{aligned}$$

Thus, the probability that $|\mathbf{pl}_j(\Lambda_{t+1}) - \mathbf{pl}_j(\Lambda_t)|$ is large is no more than the probability that $\max_{\alpha \in T_j} |Y_j(\alpha)|$ is large, and we conclude from (29) that

$$\mathbb{P} \{ |\Lambda_{t+1}|^2 (\mathbf{pl}_j(\Lambda_t) - \mathbf{pl}_j(\Lambda_{t+1}))^2 > v \} \leq 4L \exp \left(\frac{-v}{2 \mathbf{pl}_j(\Lambda_t) |\Lambda_t|} \right).$$

Integrating, we bound the expectation by

$$\begin{aligned} \mathbb{E} |\Lambda_{t+1}|^2 (\mathbf{pl}_j(\Lambda_t) - \mathbf{pl}_j(\Lambda_{t+1}))^2 &= \int_0^\infty \mathbb{P} \left\{ \max_{\alpha \in T_j} Y_j(\alpha)^2 > v \right\} dv \\ &\leq A + 4L \int_A^\infty \exp \left(\frac{-v}{2 \mathbf{pl}_j(\Lambda_t) |\Lambda_t|} \right) dv \\ &= A + 4L \cdot 2 \mathbf{pl}_j(\Lambda_t) |\Lambda_t| \cdot \exp \left(\frac{-A}{2 \mathbf{pl}_j(\Lambda_t) |\Lambda_t|} \right) \end{aligned}$$

for any $A \geq 0$. Choosing $A = 2 \mathbf{pl}_j(\Lambda_t) |\Lambda_t| \ln(4L)$ gives

$$\mathbb{E} |\Lambda_{t+1}|^2 (\mathbf{pl}_j(\Lambda_t) - \mathbf{pl}_j(\Lambda_{t+1}))^2 \leq 2 |\Lambda_t| \mathbf{pl}_j(\Lambda_t) (\ln(4L) + 1).$$

Setting C_5 correctly proves the second item in Lemma 2, and the first follows from Jensen's inequality. \square

The next lemma uses Lemma 2 to argue that a number of good things happen all at once.

Lemma 3. *There is some $\Lambda_{t+1} \subseteq \Lambda_t$ so that:*

1.

$$\left(\frac{1-\eta}{2} \right)^{t+1} L \leq \left(\frac{1-\eta}{2} \right) |\Lambda_t| \leq |\Lambda_{t+1}| \leq \left(\frac{1+\eta}{2} \right) |\Lambda_t| \leq \left(\frac{1+\eta}{2} \right)^{t+1} L.$$

2.

$$\sum_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_{t+1}) \leq \sum_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_t) + \sum_{j \in I_{t+1}} \sqrt{\frac{c_1 |\Lambda_t| \log(L) \mathbf{pl}_j(\Lambda_t)}{|\Lambda_{t+1}|^2}}$$

3.

$$\left(\sum_{j \in I_{t+1}} (\mathbf{pl}_j(\Lambda_{t+1}) - \mathbf{pl}_j(\Lambda_t))^2 \right)^{1/2} \leq \frac{\sqrt{c_1 |\Lambda_t| \log(L) Q_t}}{|\Lambda_{t+1}|}$$

for some constant c_1 .

Proof. We show that (for an appropriate choice of c_1), each of these items occurs with probability at least $2/3$, $3/4$, and $3/4$, respectively. Thus, all three occur with probability at least $1/6$, and in particular there is a set Λ_{t+1} which satisfies all three.

First, we address Item 1. By a Chernoff bound,

$$\mathbb{P} \left\{ \left| |\Lambda_{t+1}| - \frac{1}{2} |\Lambda_t| \right| > u \right\} \leq 2 \exp(-2u^2/|\Lambda_t|),$$

By the inductive hypothesis (15),

$$|\Lambda_t| \geq \left(\frac{1-\eta}{2} \right)^t L,$$

and so by our choice of t_{\max} and the fact that $t \leq t_{\max}$, we have

$$|\Lambda_t| \geq 4/\eta^2. \tag{31}$$

Thus,

$$\mathbb{P} \left\{ \left| |\Lambda_{t+1}| - \frac{|\Lambda_t|}{2} \right| \geq \frac{\eta |\Lambda_t|}{2} \right\} \leq 2e^{-2} < 1/3.$$

Again by the inductive hypothesis (15) applied to $|\Lambda_t|$, we conclude that

$$\left(\frac{1-\eta}{2} \right)^{t+1} L \leq \left(\frac{1-\eta}{2} \right) |\Lambda_t| \leq |\Lambda_{t+1}| \leq \left(\frac{1+\eta}{2} \right) |\Lambda_t| \leq \left(\frac{1+\eta}{2} \right)^{t+1} L.$$

For Item 2, we invoke Lemma 2 and linearity of expectation to obtain

$$\mathbb{E} \sum_{j \in I_{t+1}} |\Lambda_{t+1}| |\mathbf{pl}_j(\Lambda_t) - \mathbf{pl}_j(\Lambda_{t+1})| \leq \sum_{j \in I_{t+1}} \sqrt{C_5 \log(L) \mathbf{pl}_j(\Lambda_t) |\Lambda_t|}.$$

By Markov's inequality, as long as $c_1 \geq 16C_5$, with probability at least $3/4$,

$$\sum_{j \in I_{t+1}} |\Lambda_{t+1}| |\mathbf{pl}_j(\Lambda_t) - \mathbf{pl}_j(\Lambda_{t+1})| \leq \sum_{j \in I_{t+1}} \sqrt{c_1 \log(L) \mathbf{pl}_j(\Lambda_t) |\Lambda_t|},$$

and in the favorable case the triangle inequality implies

$$\begin{aligned} \sum_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_{t+1}) &\leq \sum_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_t) + \sum_{j \in I_{t+1}} |\mathbf{pl}_j(\Lambda_t) - \mathbf{pl}_j(\Lambda_{t+1})| \\ &\leq \sum_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_t) + \sum_{j \in I_{t+1}} \frac{\sqrt{c_1 \log(L) \mathbf{pl}_j(\Lambda_t) |\Lambda_t|}}{|\Lambda_{t+1}|}. \end{aligned}$$

Thus, Item 2 holds with probability at least $3/4$.

Similarly, for Item 3, Lemma 2 and linearity of expectation (as well as Jensen's inequality) implies that

$$\begin{aligned} \mathbb{E} \left(\sum_{j \in I_{t+1}} |\Lambda_{t+1}|^2 (\mathbf{pl}_j(\Lambda_{t+1}) - \mathbf{pl}_j(\Lambda_t))^2 \right)^{1/2} &\leq \left(\sum_{j \in I_{t+1}} C_5 |\Lambda_t| \log(L) \mathbf{pl}_j(\Lambda_t) \right)^{1/2} \\ &\leq \left(\sum_{j \in I_t} C_5 |\Lambda_t| \log(L) \mathbf{pl}_j(\Lambda_t) \right)^{1/2} \quad \text{since } I_{t+1} \subseteq I_t \\ &\leq \sqrt{C_5 |\Lambda_t| \log(L) Q_t} \quad \text{by the inductive hypothesis (14)}. \end{aligned}$$

Again, Markov's inequality and an appropriate restriction on c_1 implies that Item 3 occurs with probability strictly more than $3/4$.

This concludes the proof of Lemma 3. \square

Finally, we show how Lemma 3 implies the conclusions of Lemma 1 for $t + 1$, notably (14), (15) and (16). First, we observe that (15) follows immediately from Lemma 3, Item 1. Next we consider (14). The definition of I_{t+1} and the choice of γ , along with the fact from Lemma 3, Item 1 that $|\Lambda_{t+1}| \geq \left(\frac{1-\eta}{2}\right) |\Lambda_t|$, imply that for $j \in I_{t+1}$,

$$|\Lambda_t| \mathbf{pl}_j(\Lambda_t) \geq \gamma \geq \left(\frac{|\Lambda_t|}{|\Lambda_{t+1}|}\right)^2 \frac{c_1 \log(L)}{\eta^2},$$

and so

$$\frac{\sqrt{c_1 |\Lambda_t| \log(L)} \mathbf{pl}_j(\Lambda_t)}{|\Lambda_{t+1}|} \leq \eta \mathbf{pl}_j(\Lambda_t). \quad (32)$$

Thus,

$$\begin{aligned} \sum_{j \in I_{t+1}} \mathbf{pl}_j(\Lambda_{t+1}) &\leq \sum_{j \in I_{t+1}} (1 + \eta) \mathbf{pl}_j(\Lambda_t) && \text{by Lemma 3, Item 2 and from (32)} \\ &\leq (1 + \eta) \sum_{j \in I_t} \mathbf{pl}_j(\Lambda_t) && \text{since } I_{t+1} \subseteq I_t, \text{ by (28)} \\ &\leq (1 + \eta) Q_t && \text{by the inductive hypothesis (14) for } t \\ &= (1 + \eta)^{t+1} Q && \text{by the definition of } Q_t \\ &= Q_{t+1}. \end{aligned}$$

This establishes (14).

To establish the distance criterion (16), we use the triangle inequality to write

$$\|\mathbf{pl}_{I_t}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1})\|_2 = \|\mathbf{pl}_{I_{t+1}}(\Lambda_t) + \mathbf{pl}_{I_t \setminus I_{t+1}}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1})\|_2 \quad (33)$$

$$\leq \|\mathbf{pl}_{I_{t+1}}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1})\|_2 \quad (34)$$

$$+ \|\mathbf{pl}_{I_t \setminus I_{t+1}}(\Lambda_t)\|_2 \quad (35)$$

The first term (34) is bounded by Lemma 3, Item 3, by

$$\|\mathbf{pl}_{I_{t+1}}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1})\|_2 \leq \frac{\sqrt{c_1 |\Lambda_t| \log(L)} Q_t}{|\Lambda_{t+1}|}.$$

To bound (35), we will bound both the ℓ_∞ and ℓ_1 norms of $\mathbf{pl}_{I_t \setminus I_{t+1}}(\Lambda_t)$ and use Hölder's inequality to control the ℓ_2 norm. By the inductive hypothesis (14) and the fact (28) that $I_{t+1} \subseteq I_t$,

$$\|\mathbf{pl}_{I_t \setminus I_{t+1}}(\Lambda_t)\|_1 \leq \|\mathbf{pl}_{I_t}(\Lambda_t)\|_1 \leq Q_t.$$

Also, by the definition of I_{t+1} ,

$$\|\mathbf{pl}_{I_t \setminus I_{t+1}}(\Lambda_t)\|_\infty \leq \frac{\gamma}{|\Lambda_t|}.$$

Together, Hölder's inequality implies that

$$\|\mathbf{pl}_{I_t \setminus I_{t+1}}(\Lambda_t)\|_2 \leq \sqrt{\|\mathbf{pl}_{I_t \setminus I_{t+1}}(\Lambda_t)\|_1 \|\mathbf{pl}_{I_t \setminus I_{t+1}}(\Lambda_t)\|_\infty} \leq \sqrt{\frac{\gamma Q_t}{|\Lambda_t|}}.$$

This bounds the second term (35) of (33), and putting it all together we have

$$\|\mathbf{pl}_{I_t}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1})\|_2 \leq \frac{\sqrt{c_1 |\Lambda_t| \log(L)} Q_t}{|\Lambda_{t+1}|} + \sqrt{\frac{\gamma Q_t}{|\Lambda_t|}}.$$

Using the fact from Lemma 3, Item 1 that $|\Lambda_t|/|\Lambda_{t+1}| \leq 2/(1-\eta)$, as well as the definition of γ in (27), we may bound the above expression by

$$\|\mathbf{pl}_{I_t}(\Lambda_t) - \mathbf{pl}_{I_{t+1}}(\Lambda_{t+1})\|_2 \leq \left(1 + \frac{1}{\eta}\right) \left(\frac{2}{1-\eta}\right) \sqrt{\frac{c_1 \log(L) Q_t}{|\Lambda_t|}}.$$

This establishes (16), for an appropriate choice of C_4 , and for sufficiently large L (and hence sufficiently small η).

Finally, we verify the condition (17) on the size $|\mathcal{S}_{t+1}|$. By (15), and the fact that our choices of η and t_{\max} imply that $(1+\eta)^t \leq e$, $|\Lambda_t| \leq eL/2^t$. We saw earlier that I_{t+1} depends only on Λ_t , so (using the fact that $L \leq N/2$), there are at most

$$\sum_{r=1}^{eL/2^t} \binom{N}{r} \lesssim \binom{N}{eL/2^t}$$

choices for I_{t+1} . Similarly, we just chose Λ_{t+1} so that $|\Lambda_{t+1}| \leq eL/2^{t+1}$, so there are at most $\sum_{r=1}^{eL/2^{t+1}} \binom{N}{r} \lesssim \binom{N}{eL/2^{t+1}}$ choices for Λ_{t+1} . Altogether, there are at most

$$C_6 \binom{N}{eL/2^t} \binom{N}{eL/2^{t+1}}$$

choices for the pair (I_{t+1}, Λ_{t+1}) , for an appropriate constant C_6 , and this establishes (15).

This completes the proof of Lemma 1.

7 Conclusion and future work

We have shown that “most” Reed-Solomon codes are list decodable beyond the Johnson bound, answering a long-standing open question (Question 1) of [GS98, Gur04, Rud07, Vad12]. More precisely, we have shown that with high probability, a Reed-Solomon code with random evaluation points of rate

$$\Omega\left(\frac{\varepsilon}{\log(q) \log^5(1/\varepsilon)}\right)$$

is list decodable up to a $1-\varepsilon$ fraction of errors with list size $O(1/\varepsilon)$. This beats the Johnson bound whenever $\varepsilon \leq \tilde{O}(1/\log(q))$.

Our proof actually applies more generally to randomly punctured codes, and provides a positive answer to our second motivating question, Question 2, about whether randomly punctured codes with good distance can beat the Johnson bound. As an added corollary, we have obtained improved bounds on the list decodability of random linear codes over large alphabets. Our bounds are nearly optimal (up to polylogarithmic factors), and are the best known whenever $q \gtrsim \log^5(1/\varepsilon)$.

The most obvious open question that remains is to remove the polylogarithmic factors from the rate bound. The factor of $\log(q)$ is especially troublesome: it bites when $q = 2^{\Omega(1/\varepsilon)}$ is very large, but this parameter regime can be reasonable for Reed-Solomon codes. Removing this logarithmic factor seems as though it may require a restructuring of the argument. A second question is to resolve the discrepancy between our upper bound on list sizes and the bound associated with general random codes of the same rate; there is a gap of a factor of ε in the parameter regime $1/\varepsilon \leq q \leq 1/\varepsilon^2$.

To avoid ending on the shortcomings of our argument, we mention a few hopeful directions for future work. Our argument applies to randomly punctured codes in general, and it is natural to ask for more examples of codes where Theorem 2 can improve the status quo. Additionally, list decodable codes are connected to many other pseudorandom objects; it would be extremely interesting to explore the ramifications of our argument for random families of extractors or expanders, for example.

Acknowledgments

We are very grateful to Mahmoud Abo Khamis, Venkat Guruswami, Yi Li, Anindya Patthak, and Martin Strauss for careful proof-reading and helpful comments. We thank Venkat for suggesting the title of the paper. AR would like to thank Swastik Kopparty and Shubhangi Saraf for some discussions on related questions at Dagstuhl that led to questions considered in this paper.

References

- [BSKR10] Eli Ben-Sasson, Swastik Kopparty, and Jaikumar Radhakrishnan. Subspace polynomials and limits to list decoding of reed-solomon codes. *IEEE Transactions on Information Theory*, 56(1):113–120, 2010.
- [CGV13] Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of fourier matrices and list decodability of random linear codes. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 432–442, 2013.
- [CPS99] Jin-yi Cai, Aduri Pavan, and D. Sivakumar. On the hardness of permanent. In *Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 90–99, 1999.
- [CW07] Qi Cheng and Daqing Wan. On the list and bounded distance decodability of reed-solomon codes. *SIAM J. Comput.*, 37(1):195–209, 2007.
- [DL12] Zeev Dvir and Shachar Lovett. Subspace evasive sets. In *Proceedings of the 44th Symposium on Theory of Computing Conference (STOC)*, pages 351–358, 2012.
- [Eli57] Peter Elias. List decoding for noisy channels. *Technical Report 335, Research Laboratory of Electronics, MIT*, 1957.
- [GHK11] Venkatesan Guruswami, Johan Håstad, and Swastik Kopparty. On the list-decodability of random linear codes. *IEEE Transactions on Information Theory*, 57(2):718–725, 2011.
- [GK13] Venkatesan Guruswami and Swastik Kopparty. Explicit subspace designs. In *FOCS*, 2013. To appear.
- [GKZ08] Parikshit Gopalan, Adam R. Klivans, and David Zuckerman. List-decoding reed-muller codes over small fields. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 265–274, 2008.
- [GN13] Venkatesan Guruswami and Srivatsan Narayanan. Combinatorial limitations of average-radius list decoding. *RANDOM*, 2013.
- [Gop10] Parikshit Gopalan. A fourier-analytic approach to reed-muller decoding. In *Proceedings of the 51th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 685–694, 2010.
- [GR06] Venkatesan Guruswami and Atri Rudra. Limits to list decoding reed-solomon codes. *IEEE Transactions on Information Theory*, 52(8):3642–3649, 2006.
- [GR08] Venkatesan Guruswami and Atri Rudra. Explicit codes achieving list decoding capacity: Error-correction with optimal redundancy. *IEEE Transactions on Information Theory*, 54(1):135–150, 2008.

- [GS98] Venkatesan Guruswami and Madhu Sudan. Improved decoding of reed-solomon and algebraic-geometric codes. In *Proceedings of 39th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 28–39, 1998.
- [GS99] Venkatesan Guruswami and Madhu Sudan. Improved decoding of reed-solomon and algebraic-geometry codes. *IEEE Transactions on Information Theory*, 45(6):1757–1767, 1999.
- [GS01] Venkatesan Guruswami and Madhu Sudan. Extensions to the johnson bound, 2001.
- [GS03] Venkatesan Guruswami and Igor Shparlinski. Unconditional proof of tightness of johnson bound. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 754–755, 2003.
- [Gur04] Venkatesan Guruswami. *List Decoding of Error-Correcting Codes (Winning Thesis of the 2002 ACM Doctoral Dissertation Competition)*, volume 3282 of *Lecture Notes in Computer Science*. Springer, 2004.
- [GV10] Venkatesan Guruswami and Salil Vadhan. A lower bound on list size for list decoding. *Information Theory, IEEE Transactions on*, 56(11):5681–5688, 2010.
- [GW13] Venkatesan Guruswami and Carol Wang. Linear-algebraic list decoding for variants of reed-solomon codes. *IEEE Transactions on Information Theory*, 59(6):3257–3268, 2013.
- [GX12] Venkatesan Guruswami and Chaoping Xing. Folded codes from function field towers and improved optimal rate list decoding. In *Proceedings of the 44th Symposium on Theory of Computing Conference (STOC)*, pages 339–350, 2012.
- [GX13] Venkatesan Guruswami and Chaoping Xing. List decoding reed-solomon, algebraic-geometric, and gabidulin subcodes up to the singleton bound. In *Proceedings of the 45th ACM Symposium on the Theory of Computing (STOC)*, pages 843–852, 2013.
- [Kop12] Swastik Kopparty. List-decoding multiplicity codes. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:44, 2012.
- [LT91] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*, volume 23. Springer, 1991.
- [MS77] F. J. MacWilliams and N. J. A. Sloane. *The theory of error correcting codes / F.J. MacWilliams, N.J.A. Sloane*. North-Holland Pub. Co. ; sole distributors for the U.S.A. and Canada, Elsevier/North-Holland Amsterdam ; New York : New York, 1977.
- [PV05] Farzad Parvaresh and Alexander Vardy. Correcting errors beyond the guruswami-sudan radius in polynomial time. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 285–294, 2005.
- [Rud97] Mark Rudelson. Contact points of convex bodies. *Israel Journal of Mathematics*, 101(1):93–124, 1997.
- [Rud07] Atri Rudra. *List decoding and property testing of error-correcting codes*. PhD thesis, University of Washington, 2007.
- [RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.
- [Sud97] Madhu Sudan. Decoding of reed solomon codes beyond the error-correction bound. *J. Complexity*, 13(1):180–193, 1997.
- [Sud00] Madhu Sudan. List decoding: algorithms and applications. *SIGACT News*, 31(1):16–27, 2000.

- [Tal05] Michel Talagrand. *The generic chaining: upper and lower bounds for stochastic processes*. Springer, 2005.
- [Vad12] Salil P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.
- [Woo13] Mary Wootters. On the list decodability of random linear codes with large error rates. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, pages 853–860. ACM, 2013.
- [Woz58] John M. Wozencraft. List Decoding. *Quarterly Progress Report, Research Laboratory of Electronics, MIT*, 48:90–95, 1958.

A Proofs of Corollaries 1 and 2

In this appendix, we first prove a few variants on the Johnson bound, which are needed for the proofs of Corollaries 1 and 2. We require average-radius versions of two statements of the Johnson bound, found in [GS01] and [MS77], respectively. It appears to be folklore that such statements are true (and follow from the proofs in the two works cited above). For completeness, we include proofs below in Section A.1. Finally, we prove Corollaries 1 and 2 in Section A.2.

A.1 Average radius Johnson bounds

Theorem 4. *Let $\mathcal{C} : \mathbb{F}_q^k \rightarrow \mathbb{F}_q^n$ be any code. Then for all $\Lambda \subset \mathbb{F}_q^k$ of size L and for all $z \in \mathbb{F}_q^n$,*

$$\sum_{x \in \Lambda} \text{agr}(c(x), z) \leq \frac{nL}{q} + \frac{nL}{2\varepsilon} (1 + \varepsilon^2) \left(1 - \frac{1}{q}\right) - \frac{n}{2L\varepsilon} \sum_{x \neq y \in \Lambda} d(c(x), c(y)).$$

Remark 6. *An average-radius q -ary Johnson bound follows from Theorem 4 by bounding $d(c(x), c(y))$ by $1 - 1/q - \varepsilon^2$. In this case, the theorem implies that any code of distance $1 - 1/q - \varepsilon^2$ is average-radius list decodable up to error rate $\rho = 1 - 1/q - \varepsilon$ as long as the list size L obeys $L \geq 2/\varepsilon^2$.*

Proof. Fix a $z \in \mathbb{F}_q^n$. The crux of the proof is to map the relevant vectors over \mathbb{F}_q^n to vectors in \mathbb{R}^{nq} as follows. Given a vector $u \in \mathbb{F}_q^n$, let $u' \in \mathbb{R}^{nq}$ denote the concatenation

$$u' = (e_{u_1}, e_{u_2}, \dots, e_{u_n}),$$

where $e_{u_i} \in \{0, 1\}^q$ is the vector which is one in the u_i 'th index and zero elsewhere. (Above, we fix an arbitrary mapping of \mathbb{F}_q to $[q]$). In particular, for an $x \in \Lambda$, we will use $c'(x)$ to denote the mapping of the codeword $c(x)$. Finally let $v \in \mathbb{R}^{nq}$ be

$$v = \varepsilon \cdot z' + \left(\frac{1 - \varepsilon}{q}\right) \cdot \mathbf{1},$$

where $\mathbf{1}$ denotes the all-ones vector.

Given the definitions above, it can be verified that the identities below hold for every $x \neq y \in \Lambda$:

$$\langle c'(x), v \rangle = \varepsilon \cdot \text{agr}(c(x), z) + \frac{(1 - \varepsilon)n}{q}, \tag{36}$$

$$\langle v, v \rangle = \frac{n}{q} + \varepsilon^2 \left(1 - \frac{1}{q}\right) n, \tag{37}$$

$$\langle c'(x), c'(y) \rangle = n(1 - d(c(x), c(y))), \tag{38}$$

and

$$\langle c'(x), c'(x) \rangle = n. \quad (39)$$

Now consider the following sequence of relations:

$$0 \leq \left\langle \sum_{x \in \Lambda} (c'(x) - v), \sum_{x \in \Lambda} (c'(x) - v) \right\rangle \quad (40)$$

$$\begin{aligned} &= \sum_{x, y \in \Lambda} \langle c'(x), c'(y) \rangle - \sum_{x, y \in \Lambda} (\langle c'(x), v \rangle + \langle c'(y), v \rangle) + \sum_{x, y \in \Lambda} \langle v, v \rangle \\ &= \sum_{x \in \Lambda} \langle c'(x), c'(x) \rangle + \sum_{x \neq y \in \Lambda} \langle c'(x), c'(y) \rangle - 2L \cdot \sum_{x \in \Lambda} \langle c'(x), v \rangle + \sum_{x, y \in \Lambda} \langle v, v \rangle \\ &= nL + n \sum_{x \neq y \in \Lambda} (1 - d(c(x), c(y))) - 2L \cdot \sum_{x \in \Lambda} \left(\varepsilon \cdot \text{agr}(c(x), z) + \frac{(1 - \varepsilon)n}{q} \right) + L^2 \cdot \left(\frac{n}{q} + \varepsilon^2 \left(1 - \frac{1}{q} \right) n \right) \end{aligned} \quad (41)$$

$$\begin{aligned} &= nL^2 \cdot \left(1 + \frac{1}{q} + \varepsilon^2 \left(1 - \frac{1}{q} \right) - \frac{2(1 - \varepsilon)}{q} \right) - n \sum_{x \neq y \in \Lambda} d(c(x), c(y)) - 2L\varepsilon \cdot \sum_{x \in \Lambda} \text{agr}(c(x), z) \\ &= nL^2 \cdot \left((1 + \varepsilon^2) \left(1 - \frac{1}{q} \right) + \frac{2\varepsilon}{q} \right) - n \sum_{x \neq y \in \Lambda} d(c(x), c(y)) - 2L\varepsilon \cdot \sum_{x \in \Lambda} \text{agr}(c(x), z) \end{aligned} \quad (42)$$

In the above, (40) follows from the fact that the norm of a vector is always positive and (41) follows from (36), (37), (38) and (39).

Equation (42) then implies that

$$2L\varepsilon \cdot \sum_{x \in \Lambda} \text{agr}(c(x), z) \leq nL^2 \cdot \left((1 + \varepsilon^2) \left(1 - \frac{1}{q} \right) + \frac{2\varepsilon}{q} \right) - n \sum_{x \neq y \in \Lambda} d(c(x), c(y)),$$

which implies the statement after rearranging terms. \square

Next, we prove a second average-radius variant of the Johnson bound, which has been copied almost verbatim from [MS77].

Theorem 5. *Let $\mathcal{C} : \mathbb{F}_q^k \rightarrow \mathbb{F}_q^n$ be any code. Then for all $\Lambda \subset \mathbb{F}_q^k$ of size L and for all $z \in \mathbb{F}_q^n$,*

$$\sum_{x \in \Lambda} \text{agr}(c(x), z) \leq \frac{1}{2} \left(n + \sqrt{n^2 + 4n^2L(L-1) - 4n^2 \sum_{x \neq y \in \Lambda} d(c(x), c(y))} \right).$$

Proof. For every $j \in [n]$, define

$$a_j = |\{x \in \Lambda | c(x)_j = z_j\}|.$$

Note that

$$\sum_{j=1}^n a_j = \sum_{x \in \Lambda} \text{agr}(c(x), z), \quad (43)$$

and

$$\begin{aligned}
\sum_{j=1}^n \binom{a_j}{2} &= \frac{1}{2} \cdot \sum_{j=1}^n \sum_{x \neq y \in \Lambda} \mathbf{1}_{c(x)_j = z_j} \mathbf{1}_{c(y)_j = z_j} \\
&\leq \sum_{j=1}^n \sum_{x \neq y \in \Lambda} \mathbf{1}_{c(x)_j = c(y)_j} \\
&= \frac{1}{2} \cdot \sum_{x \neq y \in \Lambda} \text{agr}(c(x), c(y)) \\
&= \frac{L(L-1)n}{2} - \frac{n}{2} \sum_{x \neq y \in \Lambda} d(c(x), c(y)). \tag{44}
\end{aligned}$$

Next, note that by the Cauchy-Schwartz inequality,

$$\sum_{j=1}^n \binom{a_j}{2} = \frac{1}{2} \left(\sum_{j=1}^n a_j^2 - \sum_{j=1}^n a_j \right) \geq \frac{1}{2n} \left(\sum_{j=1}^n a_j \right)^2 - \frac{1}{2} \sum_{j=1}^n a_j.$$

Combining the above with (43) and (44) implies that

$$\left(\sum_{x \in \Lambda} \text{agr}(c(x), z) \right)^2 - n \cdot \sum_{x \in \Lambda} \text{agr}(c(x), z) - \left(n^2 L(L-1) - n^2 \sum_{x \neq y \in \Lambda} d(c(x), c(y)) \right) \leq 0,$$

which in turn implies (by the fact that the sum we care about lies in between the two roots of the quadratic equation) that

$$\sum_{x \in \Lambda} \text{agr}(c(x), z) \leq \frac{1}{2} \left(n + \sqrt{n^2 + 4n^2 L(L-1) - 4n^2 \sum_{x \neq y \in \Lambda} d(c(x), c(y))} \right),$$

which completes the proof. \square

A.2 Proofs of Corollaries 1 and 2

The proofs of both Corollaries follow essentially from the proofs of the Johnson bound in Section A.1. We use two versions of the Johnson bound, one from [GS01] which is more useful for our “small q ” regime, and another proof from [MS77] which produces better results in the “large q ” regime.

Proof of Corollary 1. Suppose that $L \geq 2/\varepsilon^2$ and that the distance of \mathcal{C}' is at least $1 - 1/q - \varepsilon^2/2$. We need an average-radius version of the Johnson bound, which we provide in Theorem 4 in Appendix A.1. By Theorem 4, for any $z \in \mathbb{F}_q^n$ and for all $\Lambda \subset \mathbb{F}_q^k$ of size L ,

$$\sum_{x \in \Lambda} \text{agr}(c(x), z) \leq \frac{nL}{q} + \frac{nL}{2\varepsilon} (1 + \varepsilon^2) \left(1 - \frac{1}{q} \right) - \frac{n}{2L\varepsilon} \sum_{x \neq y \in \Lambda} d(c(x), c(y)). \tag{45}$$

By Theorem 2, it suffices to control \mathcal{E} . Since the right hand side above does not depend on z ,

$$\begin{aligned}
\mathcal{E} &= \max_{|\Lambda|=L} \mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^k} \sum_{x \in \Lambda} \text{agr}(c(x), z) \\
&\leq \max_{|\Lambda|=L} \mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^k} \left(\frac{nL}{q} + \frac{nL}{2\varepsilon} (1 + \varepsilon^2) \left(1 - \frac{1}{q} \right) - \frac{n}{2L\varepsilon} \sum_{x \neq y \in \Lambda} d(c(x), c(y)) \right)
\end{aligned}$$

$$\begin{aligned}
&= \max_{|\Lambda|=L} \left(\frac{nL}{q} + \frac{nL}{2\varepsilon} (1 + \varepsilon^2) \left(1 - \frac{1}{q}\right) - \frac{n}{2L\varepsilon} \sum_{x \neq y \in \Lambda} \mathbb{E}_{\mathcal{C}} d(c(x), c(y)) \right) \\
&\leq \frac{nL}{q} + \frac{nL}{2\varepsilon} (1 + \varepsilon^2) \left(1 - \frac{1}{q}\right) - \frac{n(L-1) \left(1 - \frac{1}{q} - \frac{\varepsilon^2}{2}\right)}{2\varepsilon} \tag{46}
\end{aligned}$$

$$\begin{aligned}
&= \frac{nL}{q} + \frac{nL\varepsilon}{2} \left(\frac{3}{2} - \frac{1}{q}\right) + \frac{n \left(1 - \frac{1}{q} - \frac{\varepsilon^2}{2}\right)}{2\varepsilon} \\
&\leq \frac{nL}{q} + \frac{3nL\varepsilon}{4} + \frac{n}{2\varepsilon} \\
&\leq nL \left(\frac{1}{q} + \varepsilon\right). \tag{47}
\end{aligned}$$

In the above, (46) follows from the fact that the original code had (relative) distance $1 - 1/q - \varepsilon^2/2$ and that in the construction of \mathcal{C} from \mathcal{C}' , pairwise Hamming distances are preserved in expectation. Finally, (47) follows from the assumption that $L \geq 2/\varepsilon^2$.

Recall from the statement of Theorem 2 that we have defined

$$Y = C_0 L \log(N) \log^5(L),$$

so the assumption on n implies that

$$Y \leq nL \min\{\varepsilon, q\varepsilon^2\}.$$

Suppose that $q\varepsilon \leq 1$, so that $Y \leq nLq\varepsilon^2$. Plugging this along with (47) into Theorem 2, we obtain

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^n} \max_{\Lambda \subset \mathbb{F}_q^k, |\Lambda|=L} \sum_{x \in \Lambda} \text{agr}(c(x), z) &\leq \mathcal{E} + Y + \sqrt{\mathcal{E}Y} \\
&\leq nL \left(\frac{1}{q} + \varepsilon\right) + nLq\varepsilon^2 + nL \sqrt{q\varepsilon^2 \left(\frac{1}{q} + \varepsilon\right)} \\
&= nL \left(\frac{1}{q} + \varepsilon \left(1 + q\varepsilon + \sqrt{1 + q\varepsilon}\right)\right) \\
&\leq nL \left(\frac{1}{q} + \varepsilon \left(2 + \sqrt{2}\right)\right),
\end{aligned}$$

using the assumption that $q\varepsilon \leq 1$ in the final line. Thus, Proposition 1 implies that \mathcal{C} is $(1 - 1/q - (2 + \sqrt{2})\varepsilon, 2/\varepsilon^2)$ -list-decodable.

On the other hand, suppose that $q\varepsilon \geq 1$, so that $Y \leq nL\varepsilon$. Then following the same outline, we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^n} \max_{\Lambda \subset \mathbb{F}_q^k, |\Lambda|=L} \sum_{x \in \Lambda} \text{agr}(c(x), z) &\leq \mathcal{E} + Y + \sqrt{\mathcal{E}Y} \\
&\leq nL \left(\frac{1}{q} + \varepsilon\right) + nL\varepsilon + nL \sqrt{\varepsilon \left(\frac{1}{q} + \varepsilon\right)} \\
&= nL \left(\frac{1}{q} + \varepsilon \left(2 + \sqrt{\frac{1}{q\varepsilon} + 1}\right)\right) \\
&\leq nL \left(\frac{1}{q} + \varepsilon \left(2 + \sqrt{2}\right)\right),
\end{aligned}$$

using the assumption that $q\varepsilon \geq 1$ in the final line. Thus, in this case as well, \mathcal{C} is $(1 - 1/q - (2 + \sqrt{2})\varepsilon, 2/\varepsilon^2)$ -list-decodable.

This completes the proof of Corollary 1. \square

Proof of Corollary 2. As with Corollary 1, we need an average-radius version of the Johnson bound. In this case, we follow a proof of the Johnson bound from [MS77], which gives a better dependence on ε in the list size when q is large. For completeness, our average-radius version of the proof is given in Appendix A.1, Theorem 5.

We proceed with the proof of Corollary 2. By Theorem 5, for any $z \in \mathbb{F}_q^n$ and for all $\Lambda \subset \mathbb{F}_q^k$ of size L ,

$$\left[\sum_{x \in \Lambda} \text{agr}(c(x), z) \leq \frac{1}{2} \left(n + \sqrt{n^2 + 4n^2 L(L-1) - 4n^2 \sum_{x \neq y \in \Lambda} d(c(x), c(y))} \right) \right]. \quad (48)$$

By Theorem 2, it suffices to control \mathcal{E} . Since the right hand side above does not depend on z ,

$$\begin{aligned} \mathcal{E} &= \max_{|\Lambda|=L} \mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^k} \sum_{x \in \Lambda} \text{agr}(c(x), z) \\ &\leq \max_{|\Lambda|=L} \mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^k} \left(\frac{1}{2} \left(n + \sqrt{n^2 + 4n^2 L(L-1) - 4n^2 \sum_{x \neq y \in \Lambda} d(c(x), c(y))} \right) \right) \end{aligned} \quad (49)$$

$$\leq \max_{|\Lambda|=L} \frac{1}{2} \left(n + \sqrt{n^2 + 4n^2 L(L-1) - 4n^2 \sum_{x \neq y \in \Lambda} \mathbb{E}_{\mathcal{C}} d(c(x), c(y))} \right) \quad (50)$$

$$\leq \frac{1}{2} \left(n + \sqrt{n^2 + 4n^2 L(L-1) - 4n^2 \sum_{x \neq y \in \Lambda} (1 - \varepsilon^2)} \right) \quad (51)$$

$$\leq \frac{1}{2} \left(n + \sqrt{n^2 + 4n^2 L(L-1) \varepsilon^2} \right)$$

$$< \frac{1}{2} \left(n + \sqrt{n^2 + 4n^2 L^2 \varepsilon^2} \right)$$

$$\leq 2nL\varepsilon. \quad (52)$$

In the above, (49) follows from (48). (50) follows from Jensen's inequality. (51) follows from the fact that the original code had (relative) distance $1 - \varepsilon^2$ and that in the construction of \mathcal{C} from \mathcal{C}' , pairwise Hamming distances are preserved in expectation. Finally, (52) follows from the assumption that $L \geq 1/\varepsilon$.

Now, Theorem 2 implies that

$$\begin{aligned} \mathbb{E}_{\mathcal{C}} \max_{z \in \mathbb{F}_q^n} \max_{\Lambda \subset \mathbb{F}_q^k, |\Lambda|=L} \sum_{x \in \Lambda} \text{agr}(c(x), z) &\leq \mathcal{E} + Y + \sqrt{\mathcal{E}Y} \\ &\leq 2(\mathcal{E} + Y) \\ &\leq 2(2nL\varepsilon + Y) \\ &\leq 5nL\varepsilon \end{aligned}$$

where as before

$$Y = C_0 L \log(N) \log^5(L)$$

and where we used the choice of n in the final line. Choose $\varepsilon' = 5\varepsilon$, so that whenever $5\varepsilon > 1/q$, Proposition 1 applies and completes the proof. Because we have chosen $\varepsilon > 1/\sqrt{q}$ (which is necessary in order for \mathcal{C}' to have distance $1 - \varepsilon^2$), the condition that $5\varepsilon > 1/q$ holds for sufficiently small ε . \square

B Background on Gaussian random variables

In this appendix, we record a few useful facts about Gaussian random variables which we use in the body of the paper. Next, we justify the claim (22) from Section 6.2. Finally, we justify the claim (9) from Section 5. These facts are standard, and can be found, for example, in [LT91].

B.1 Some facts about Gaussians

A gaussian random variable $g \sim N(0, \sigma^2)$ with variance σ^2 has a probability density function

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-t^2/2\sigma^2).$$

The cumulative distribution function,

$$\mathbb{P}\{g > t\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{u=t}^{\infty} \exp(-u^2/2\sigma^2) du$$

obeys the estimate

$$\mathbb{P}\{g > t\} \leq \frac{\sigma}{t} \cdot \frac{1}{\sqrt{2\pi}} \exp(-t^2/2\sigma^2) \quad (53)$$

for all $t > 0$. Indeed, because on the domain $u \geq t$, $(u/t) \geq 1$, we have

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{u=t}^{\infty} \exp\left(\frac{-u^2}{2\sigma^2}\right) du \leq \frac{1}{\sqrt{2\pi\sigma^2}} \int_{u=t}^{\infty} \frac{u}{t} \exp\left(\frac{-u^2}{2\sigma^2}\right) du = \frac{\sigma}{t\sqrt{2\pi}} \exp\left(\frac{-t^2}{2\sigma^2}\right). \quad (54)$$

Linear combinations of Gaussian random variables are again Gaussian.

Fact 6. *Let g_1, \dots, g_n be Gaussian random variables with variances $\sigma_1^2, \dots, \sigma_n^2$. Then the random variable $\sum_i a_i g_i$ is again a Gaussian random variable, with variance $\sum a_i^2 \sigma_i^2$.*

In the body of the paper, we use the following bound on the expected value of the maximum of n Gaussian random variables with variances $\sigma_1^2, \dots, \sigma_n^2 \leq \sigma^2$.

Proposition 2. *Let $g_i \sim N(0, \sigma_i^2)$, for $i = 1, \dots, n$, and suppose that $\max_i \sigma_i \leq \sigma$. Then*

$$\mathbb{E} \max_{i \in [n]} |g_i| \leq \sigma \sqrt{2 \ln(n)} \cdot (1 + o(1)).$$

Proof. We have

$$\begin{aligned} \mathbb{E} \max_{i \in [n]} |g_i| &= \int_{u=0}^{\infty} \mathbb{P} \left\{ \max_{i \in [n]} |g_i| > u \right\} du \\ &\leq A + \frac{2}{\sqrt{2\pi}} \int_{u=A}^{\infty} n \exp\left(\frac{-u^2}{2\sigma^2}\right) du \end{aligned}$$

for any $A \geq \sigma$ (which we will choose shortly). In the above inequality, we have used (53) (with the fact that $A \geq \sigma$) and the fact that for every i , $\mathbb{P}\{|g_i| > u\} = 2\mathbb{P}\{g_i > u\}$. We may estimate the integral using (54), so

$$\frac{2}{\sqrt{2\pi}} \int_{u=A}^{\infty} \exp\left(\frac{-u^2}{2\sigma^2}\right) du \leq \frac{2\sigma^2}{A\sqrt{2\pi}} \exp\left(-\frac{A^2}{2\sigma^2}\right).$$

Choosing $A = \sigma\sqrt{2 \ln(n)}$, we get

$$\mathbb{E} \max_{i \in [n]} |g_i| \leq \sigma \sqrt{2 \ln(n)} + \frac{\sigma}{\sqrt{\pi \ln(n)}}.$$

□

B.2 Justification of (22)

We use a computation similar to that in the proof of Proposition 2 to justify (22), which states that

$$\mathbb{E} \max_{(I,\Lambda) \in \mathcal{S}_0} S(I, \Lambda) \leq 2 \sum_{t=1}^{t_{\max}-1} a_t =: 2A.$$

Recall that we had shown that

$$\mathbb{P} \left\{ \max_{(I,\Lambda) \in \mathcal{S}_0} S(I, \Lambda) > u \cdot \sum_{t=0}^{t_{\max}-1} a_t \right\} \leq \sum_{t=0}^{t_{\max}-1} N_t N_{t+1} \exp \left(-\frac{u^2 \cdot a_t^2}{\delta_t^2} \right),$$

and that we had chosen

$$a_t = \sqrt{2 \ln(N_t N_{t+1})} \delta_t.$$

Now (22) follows from a computation similar to the proof of Proposition 2. Indeed, we have

$$\begin{aligned} \mathbb{E} \max_{(I,\Lambda) \in \mathcal{S}_0} S(I, \Lambda) &= \int_{u=0}^{\infty} \mathbb{P} \left\{ \max_{(I,\Lambda)} S(I, \Lambda) > u \right\} du \\ &\leq A + \int_{u=A}^{\infty} \sum_{t=0}^{t_{\max}-1} N_t N_{t+1} \exp \left(\frac{-u^2 \cdot a_t^2}{\delta_t^2 A^2} \right) du \\ &= A + \int_{u=A}^{\infty} \sum_{t=0}^{t_{\max}-1} N_t N_{t+1} \exp \left(\frac{-2u^2 \ln(N_t N_{t+1})}{A^2} \right) du \\ &\leq A + \sum_{t=0}^{t_{\max}-1} N_t N_{t+1} \int_{u=A}^{\infty} \exp \left(\frac{-2u^2 \ln(N_t N_{t+1})}{A^2} \right) du. \end{aligned}$$

Repeating the trick (54), we estimate

$$\int_{u=A}^{\infty} \exp \left(\frac{-2u^2 \ln(N_t N_{t+1})}{A^2} \right) du \leq \frac{A}{4 \ln(N_t N_{t+1})} \exp(-2 \ln(N_t N_{t+1})) \leq \frac{A}{4 N_t^2 N_{t+1}^2}.$$

Plugging this in, we get

$$\mathbb{E} \max_{(I,\Lambda) \in \mathcal{S}_0} S(I, \Lambda) \leq A \left(1 + \frac{1}{4} \sum_{t=0}^{t_{\max}-1} \frac{1}{N_t N_{t+1}} \right) \leq 2A.$$

In the last inequality, we used the definition of $N_t = C_6 \binom{N}{\epsilon_L/2^t} \binom{N}{\epsilon_L/2^{t+1}}$ if $t \geq 1$ and $N_0 = \binom{N}{L}$. In particular, we have used the fact that $N_t \geq 2$ for our setting of parameters.

B.3 Justification of (9)

Finally, we justify (9), which read

$$\mathbb{E}_{\mathcal{C}} \max_{|\Lambda|=L} \left| \sum_{j \in [n]} (\mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}} \mathbf{pl}_j(\Lambda)) \right| \leq \sqrt{2\pi} \mathbb{E}_{\mathcal{C}} \mathbb{E}_g \max_{|\Lambda|=L} \left| \sum_{j \in [n]} g_j \mathbf{pl}_j(\Lambda) \right|.$$

Recall that the $\mathbf{pl}_j(\Lambda)$ are independent random variables. We proceed in two steps; first, a symmetrization argument will introduce Rademacher random variables⁹ ξ_i , and next a comparison argument will replace these with Gaussian random variables. Both steps are standard, and more general versions are given in [LT91] as Lemma 6.3 and Equation (4.8), respectively. Here, we state and prove simplified versions for our needs.

We begin by symmetrizing the left hand side of (9).

⁹That is, random variables which take the values $+1$ and -1 with probability $1/2$ each.

Lemma 4. With $\mathbf{pl}_j(\Lambda)$ as above,

$$\mathbb{E} \max_{|\Lambda|=L} \left| \sum_{j \in [n]} \mathbf{pl}_j(\Lambda) - \mathbb{E} \mathbf{pl}_j(\Lambda) \right| \leq 2 \mathbb{E} \max_{|\Lambda|=L} \left| \sum_{j \in [n]} \xi_j \mathbf{pl}_j(\Lambda) \right|,$$

where the ξ_i are independent Rademacher random variables.

Proof. Let \mathcal{C}' be an independent copy of \mathcal{C} , and let $\mathbf{pl}'_j(\Lambda)$ denote an independent copy of $\mathbf{pl}_j(\Lambda)$. Then,

$$\begin{aligned} \mathbb{E}_{\mathcal{C}} \max_{\Lambda} \left| \sum_{j \in [n]} \mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}} \mathbf{pl}_j(\Lambda) \right| &= \mathbb{E}_{\mathcal{C}} \max_{\Lambda} \left| \sum_{j \in [n]} \mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}} \mathbf{pl}_j(\Lambda) - \mathbb{E}_{\mathcal{C}'} [\mathbf{pl}'_j(\Lambda) - \mathbb{E}_{\mathcal{C}'} \mathbf{pl}'_j(\Lambda)] \right| \\ &\leq \mathbb{E}_{\mathcal{C}} \mathbb{E}_{\mathcal{C}'} \max_{\Lambda} \left| \sum_{j \in [n]} \mathbf{pl}_j(\Lambda) - \mathbf{pl}'_j(\Lambda) \right| && \text{by Jensen's inequality,} \\ &&& \text{and because } \mathbb{E}_{\mathcal{C}} \mathbf{pl}_j(\Lambda) = \mathbb{E}_{\mathcal{C}'} \mathbf{pl}'_j(\Lambda) \\ &= \mathbb{E}_{\xi} \mathbb{E}_{\mathcal{C}} \mathbb{E}_{\mathcal{C}'} \max_{\Lambda} \left| \sum_{j \in [n]} \xi_j (\mathbf{pl}_j(\Lambda) - \mathbf{pl}'_j(\Lambda)) \right| && \text{by independence, and} \\ &&& \text{the fact that } \mathbf{pl}_j(\Lambda) \text{ and } \mathbf{pl}'_j(\Lambda) \\ &&& \text{are identically distributed} \\ &\leq 2 \mathbb{E}_{\xi} \mathbb{E}_{\mathcal{C}} \max_{\Lambda} \left| \sum_{j \in [n]} \xi_j \mathbf{pl}_j(\Lambda) \right| && \text{by the triangle inequality.} \end{aligned}$$

□

Next, we replace the Rademacher random variables ξ_j with Gaussian random variables g_j using a comparison argument.

Lemma 5. Condition on the choice of \mathcal{C} , and let $\mathbf{pl}_j(\Lambda)$ be as above. Let ξ_1, \dots, ξ_n be independent Rademacher random variables, and let g_1, \dots, g_n be independent standard normal random variables. Then

$$\mathbb{E}_{\xi} \max_{|\Lambda|=L} \left| \sum_{j \in [n]} \xi_j \mathbf{pl}_j(\Lambda) \right| \leq \sqrt{\frac{\pi}{2}} \mathbb{E}_g \max_{|\Lambda|=L} \left| \sum_{j \in [n]} g_j \mathbf{pl}_j(\Lambda) \right|.$$

Proof. We have

$$\begin{aligned} \mathbb{E}_g \max_{\Lambda} \left| \sum_{j \in [n]} g_j \mathbf{pl}_j(\Lambda) \right| &= \mathbb{E}_g \mathbb{E}_{\xi} \max_{\Lambda} \left| \sum_{j \in [n]} \xi_j |g_j| \mathbf{pl}_j(\Lambda) \right| \\ &\geq \mathbb{E}_{\xi} \max_{\Lambda} \left| \sum_{j \in [n]} \xi_j \mathbb{E}_g |g_j| \mathbf{pl}_j(\Lambda) \right| && \text{by Jensen's inequality} \\ &= \mathbb{E}_{\xi} \max_{\Lambda} \left| \sum_{j \in [n]} \xi_j \sqrt{\frac{2}{\pi}} \mathbf{pl}_j(\Lambda) \right|. \end{aligned}$$

Above, we used the fact that for a standard normal random variable g_j , $\mathbb{E}|g_j| = \sqrt{2/\pi}$. □

Together, Lemma 4 and Lemma 5 imply (9).