Electronic Colloquium on Computational Complexity, Report No. 17 (2012)

Combinatorial limitations of a strong form of list decoding

Venkatesan Guruswami* Srivatsan Narayanan*

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213.

Abstract

We prove the following results concerning the combinatorics of list decoding, motivated by the exponential gap between the known upper bound (of $O(1/\gamma)$) and lower bound (of $\Omega_p(\log(1/\gamma))$) for the list-size needed to decode up to radius p with rate γ away from capacity, i.e., $1 - h(p) - \gamma$ (here $p \in (0, 1/2)$ and $\gamma > 0$).

- We prove that in any binary code $C \subseteq \{0,1\}^n$ of rate $1 - h(p) - \gamma$, there must exist a set $\mathcal{L} \subset C$ of $\Omega_p(1/\sqrt{\gamma})$ codewords such that the average distance of the points in $\mathcal{L}$ from their centroid is at most pn. In other words, there must exist $\Omega_p(1/\sqrt{\gamma})$ codewords with low "average radius". The motivation for this result is that it gives a list-size lower bound for a strong notion of list decoding; this strong form has been implicitly been used in the previous negative results for list decoding. (The usual notion of list decoding corresponds to replacing *average* radius by the *minimum* radius of an enclosing Hamming ball.)

The remaining results are for the usual notion of list decoding:

- We give a short simple proof, over all fixed alphabets, of the above-mentioned $\Omega_p(\log(1/\gamma))$ lower bound due to Blinovsky.
- We show that one *cannot* improve the $\Omega_p(\log(1/\gamma))$ lower bound via techniques based on identifying the zero-rate regime for list decoding of constant-weight codes (this is a typical approach for negative results in coding theory, including the $\Omega_p(\log(1/\gamma))$ list size lower bound). On a positive note, our $\Omega_p(1/\sqrt{\gamma})$ lower bound for the strong form of list decoding does circumvent this barrier.
- We show a "reverse connection" showing that constant-weight codes for list decoding imply general codes for list decoding with higher rate. This shows that the best possible list-size, as a function of the gap γ of the rate to the capacity limit, is the same up to constant factors for both constant-weight codes and general codes.
- We give simple second moment based proofs that w.h.p. a list-size of $\Omega_p(1/\gamma)$ is needed for list decoding *random* codes from errors as well as erasures, at rates which are γ away from the corresponding capacities. For *random linear* codes, the corresponding list size bounds are $\Omega_p(1/\gamma)$ for errors and $\exp(\Omega_p(1/\gamma))$ for erasures.

*Research supported in part by NSF grants CCF 0953155 and CCF 0963975. Email: guruswami@cmu.edu, srivatsa@cs.cmu.edu

ISSN 1433-8092

1 Introduction

The list decoding problem for an error-correcting code $C \subseteq \Sigma^n$ consists of finding the set of all codeword of C with Hamming distance pn from an input string $y \in \Sigma^n$. Though it was originally introduced in early work of Elias and Wozencraft [5, 14] in the context of average decoding error probability estimation for random error models, recently the main interest in list decoding has been for adversarial error models. List decoding enables correcting up to a factor two more worst-case errors compared to algorithms that are always restricted to output a unique answer, and this potential has even been realized algorithmically [9, 7, 11].

In this work, we are interested in some fundamental combinatorial questions concerning list decoding, which highlight the important trade-offs in this model. Fix $p \in (0, 1/2)$ and a positive integer L. We say that a binary code $C \subseteq \{0,1\}^n$ is (p, L) list decodable if every Hamming ball of radius pn has less than L codewords.[1] Here, p corresponds to the error-fraction and L to the list-size needed by the error-correction algorithm. Note that (p, L)-list decodability imposes a sparsity requirement on the distribution of codewords in Hamming space. A natural combinatorial question that arises in this context is to place bounds on the largest size of a code meeting this requirement. In particular, an outstanding open question is to characterize the maximum rate (defined to be the limiting ratio $\frac{\log |C|}{n}$ as $n \to \infty$) of a (p, L)-list decodable code.

By a simple volume packing argument, it can be shown that a (p, L)-list decodable code has rate at most $1 - h(p) + o(1)$. (Throughout, for $x \in [0, 1/2]$, we use $h(x)$ to denote the binary entropy function at x.) Indeed, picking a random center x, the Hamming ball $\mathbf{B}(x, pn)$ contains at least $|C| \cdot \binom{n}{pn} 2^{-n} = |C| \cdot 2^{-(1-h(p)+o(1))n}$ in expectation. Bounding this by $(L-1)$, we get the claim. On the positive side, in the limit of large L, the rate of a (p, L)-list decodable code approaches the optimal $1 - h(p)$. More precisely, for any $\gamma > 0$, there exists a $(p, 1/\gamma)$-list decodable code of rate at least $1 - h(p) - \gamma$. In fact, a random code of rate $1 - h(p) - \gamma$ is $(p, 1/\gamma)$-list decodable whp [15, 6], [2], and a similar result holds for random linear codes (with list-size C_p/γ) [8]. In other words, a dense random packing of $2^{(1-h(p)-\gamma)n}$ Hamming balls of radius pn (and therefore volume $\approx 2^{h(p)n}$ each) is "near-perfect" whp in the sense that no point is covered by more than $O(1/\gamma)$ balls.

The determination of the best asymptotic code rate of binary (p, L)-list decodable codes as p, L are held fixed and the block length grows is wide open for every choice of $p \in (0, 1/2)$ and integer $L \geq 1$. However, we *do* know that this rate tends to $1 - h(p)$ in the limit of large $L \to \infty$. To understand this rate of convergence as a function of list size L, following [8], let us define $L_{p,\gamma}$ to be the minimum integer L such that there exist (p, L)-list decodable codes of rate $1 - h(p) - \gamma$ for infinitely many block lengths n (the quantity γ is the "gap" to "list decoding capacity"). In [1], Blinovsky showed that a (p, L)-list decodable code has rate at most $1 - h(p) - 2^{-\Theta_p(L)}$. In particular, this implies that for any $L < \infty$, a (p, L)-list decodable code has rate strictly below the optimal $1 - h(p)$. Stated in terms of $L_{p,\gamma}$, his result gives $L_{p,\gamma} \geq \Omega_p(\log(1/\gamma))$. We provide a short and simple proof of this lower bound in Section 4, which also works almost as easily over non-binary alphabets. In contrast, Blinovsky's subsequent proof for the non-binary case involved substantial technical effort [3, 4].

Observe the exponential gap (in terms of the dependence on γ) between the $O(1/\gamma)$ upper bound and $\Omega_p(\log(1/\gamma))$ lower bounds on the quantity $L_{p,\gamma}$. Despite being a basic and fundamental question about sphere packings in Hamming space and its direct relevance to list decoding,

[1] This differs from the traditional definition of (p, L)-list decodability, which require at most L codewords. The modified definition ends up being more convenient for our purposes in this paper. Further, we are interested in the regime of large L where the two definitions are almost equivalent.

[2] By using random coding with expurgation, the list size can be improved to $h(p)/\gamma$.

there has been no progress on narrowing this asymptotic gap in the 25 years since the works of Zyablov-Pinsker [15] and Blinovsky [1]. This is the motivating challenge driving this work.

1.1 Prior work on list-size lower bounds

We now discuss some lower bounds (besides Blinovsky's general lower bound) on list-size that have been obtained in restricted cases.

Rudra shows that the $O_p(1/\gamma)$ bound obtained via the probabilistic method for random codes is, in fact, tight up to constant factors [13]. Formally, there exists $L = \Omega_p(1/\gamma)$ such that a random code of rate $1 - h(p) - \gamma$ is *not* (p, L)-list decodable w.h.p. His proof uses near-capacity-achieving codes for the binary symmetric channel, the existence of which is promised by Shannon's theorem, followed by a second moment argument. We give a simpler proof via a more direct use of the second moment method. This has the advantage that it works uniformly for random general as well as random linear codes, and for channels that introduce errors as well as erasures.

Guruswami and Vadhan [10] consider the problem of list size tradeoff when the channel may corrupt close to half the bits, that is, when $p = 1/2 - \epsilon$, and more generally $p = 1 - 1/q - \varepsilon$ for codes over an alphabet of size q. (Note that decoding is impossible if the channel could corrupt up to $1/2$ fraction of bits.) They show that there exists $c > 0$ such that for all $\epsilon > 0$ and all block lengths n, any $(1/2 - \epsilon, c/\epsilon^2)$-list decodable code contains $O_\epsilon(1)$ codewords. For p bounded away from $1/2$ (or $1 - 1/q$ in the q-ary case), their methods do not yield any non-trivial list-size lower bound as a function of gap γ to list decoding capacity.

1.2 Our main results

We have already mentioned our new proofs of Blinovsky's lower bound for general codes, and the asymptotically optimal list-size lower bound for random (and random linear) codes.

Our main results are motivated by the above-mentioned approaches, based on a strong form of list decoding, used in [1, 10] to establish list-size lower bounds. In this work, we formally define the notion of (p, L)-*strong* list decodability of a code underlying these proofs. This notion is a very natural one: a code is (p, L)-*strongly list decodable* if for every L codewords, the *average* distance of their centroid from the L codewords exceeds pn. Note that this is a stronger requirement than (p, L)-list decodability where only the *maximum* distance from any center point to the L codewords must exceed pn.

We are able to prove nearly tight bounds on the achievable rate of a (p, L)-strong list decodable code. To state our result formally, denote by $L^{\mathsf{strong}}_{p,\gamma}$ the minimum L such that there exists a (p, L)-strongly list decodable code family of rate $1 - h(p) - \gamma$. A simple random coding argument shows that a random code of $1 - h(p) - \gamma$ is $(p, 1/\gamma)$-strongly list decodable (matching the list decodability of random codes). That is, $L^{\mathsf{strong}}_{p,\gamma} \leq 1/\gamma$. Our main technical result is a lower bound on the list size that is polynomially related to the upper bound, namely $L^{\mathsf{strong}}_{p,\gamma} \geq \Omega_p(\gamma^{-1/2})$.

1.3 Our other results

We also make several clarifying observations on the landscape of the bounds for list-decodable codes, as well as the general methodology of proving combinatorial limitations of list-decodable codes. Many negative results in coding theory (i.e., results which place an upper bound on rate) proceed via a typical approach in which they pass to a constant weight $\lambda \in (p, 1/2]$; that is, restrict

the codewords to be of weight exactly λn. They show that under this restriction, a code with the stated properties must have a constant number of codewords (that is, *zero* rate). Mapping this bound back to the unrestricted setting one gets a rate upper bound of $1 - h(\lambda)$ for the original problem. For instance, the Elias-Bassalygo bound for rate R vs. relative distance δ is of this nature (here λ is picked to be the Johnson radius for list decoding for codes of relative distance δ).

The above is also the approach taken in Blinovsky's work [1] as well as that of [10]. We show that such an approach does not and *cannot* give any bound better than Blinovsky's $\Omega_p(\log(1/\gamma))$ bound for $L_{p,\gamma}$. More precisely, for any $\lambda \geq p + 2^{-c_p L}$ for some $c_p > 0$, we show that there exists a (p, L)-(strongly) list decodable code of rate $\Omega_{p,L}(1)$. Thus in order to improve the lower bound, we *must* be able to handle codes of strictly positive rate, and cannot deduce the bound by pinning down the zero-rate regime of constant-weight codes. This perhaps points to why improvements to Blinovsky's bounds have been difficult. On a positive note, we remark that we *are* able to effect such a proof for strong list decodability (some details follow next).[3]

To describe the method underlying our list-size lower bound for strongly list-decodable codes, it is convenient to express the statement as an upper bound on rate in terms of list-size L. Note that a list-size lower bound of $L \geq \Omega_p(1/\sqrt{\gamma})$ for (p, L)-strongly list-decodable codes of rate $1-h(p)-\gamma$ amounts to proving an upper bound of $1 - h(p) - \Omega_p(1/L^2)$ on the rate of (p, L)-strongly list decodable codes. Our proof of such an upper bound proceeds by first showing a rate upper bound of $h(\lambda) - h(p) - \Omega_p(1/L^2)$ for such codes whose codewords are restricted to all have weight λn (for a suitable choice of $\lambda \in (p, 1/2]$). To map this back to the original setting (with no weight restrictions on codewords), one simply notes that every (p, L)-strongly list decodable code of rate R has a constant λ-weight subcode of rate $R - (1 - h(\lambda))$.

Generally speaking, by passing to a constant-weight subcode, one can translate combinatorial results on limitations of constant-weight codes to results showing limitations for the case of general codes. We are not aware of a reverse connection (for any of the standard combinatorial coding problems) that allows one to translate limitations for general codes into corresponding limitations for constant-weight codes. This leaves open the possibility that the problem of showing limitations of constant-weight codes may be harder than the corresponding problem for general codes, or worse still, have a different answer making it impossible to solve the problem for general codes via the methodology of passing to constant-weight codes.

We show that for the problem of list decoding this is fortunately not the case, and there is in fact a reverse connection of the above form. Formally, we prove that a rate upper bound of $1-h(p)-\gamma_{p,L}$ for (p, L)-list decodable codes implies a rate upper bound of $h(\lambda)-h(p)-\gamma_{p,L}\left(\frac{\lambda-p}{1/2-p}\right)$ for (p, L)-list decodable codes whose codewords must all have Hamming weight λn. A similar claim holds also for strong list decodability, though we don't state it formally.

1.4 Our proof techniques

Our proofs in this paper employ variants of the standard probabilistic method. We show an extremely simple probabilistic argument that yields a $\Omega_p(\log(1/\gamma))$ bound on the list size of a standard list decodable code; we emphasize that this is qualitatively the tightest known bound. For the "strong list decoding" problem that we introduce, we are able to improve this list-size bound to $\Omega_p(1/\sqrt{\gamma})$. The proof is based on the idea that instead of picking the "bad list decoding center"

[3] Though the technical details are very different, it may be worth noting the similarity of this with bounds for rate vs. distance. Passing to the zero-rate regime for constant-weight codes gives the Elias-Bassalygo bound, and the more sophisticated and stronger second linear programming bound is obtained by working in the regime of positive rate.

uniformly at random, one can try to pick it randomly very close to a codeword, and this still gives similar guarantees on the number of near-by codewords. Now since the quantity of interest is the average radius, this close-by codeword gives enough savings for us.

For bounds on random codes, our main novelty is to define a random variable $\mathcal{Z}$ that counts the number of "violations" of the list-decoding property of the code. We then show that $\mathcal{Z}$ has a exponentially large mean around which it is concentrated w.h.p. This yields that the code cannot be list-decodable with high probability, for suitable values of rate and list size parameters.

1.5 Organization

We define some useful notation and the formal notion of strong list decodability in Section 2. Our main negative result on limitations of strongly list-decodable codes appears in Section 3; for ease of readability, the most technical part of the proof is isolated as Appendix **??**. We give our short proof of Blinovsky's lower bound in Section 4. Our results about the zero-error rate regime for constant-weight codes and the reverse connection between list decoding bounds for general codes and constant-weight codes appear in Section 5. Finally, our list size lower bounds for random codes are discussed in Section 6, with the case of list decoding from erasures appearing as Appendix B.

2 Notation and Preliminaries

We recall some standard terminology regarding error-correcting codes. For $q \geq 2$, let $[q]$ denote the set $\{0, 1, \ldots, q-1\}$. By a q-ary code, we mean any set $C \subseteq [q]^n$, where n is called the blocklength of C. We will mainly focus on the special case of binary codes corresponding to $q = 2$. The rate $R = R(C)$ is defined to be $\frac{\log |C|}{n \log q}$.[4] For $x \in [q]^n$ and $S \subseteq [n]$, we denote by $x|_S$ the restriction of x to the coordinates in S. Let $\mathrm{supp}(x) := \{i \in [n] \,:\, x_i \neq 0\}$. A subcode of C is simply any $C' \subseteq C$.

For $x, x' \in [q]^n$, define the *Hamming distance* between x and x', denoted $d(x, x')$, to be the number of coordinates in which x and x' differ. The *weight* (or *density*) of $x \in [q]^n$, denoted $\mathrm{wt}(x)$, is $d(\mathbf{0}, x)$, where $\mathbf{0}$ is the all-zeros vector in $[q]^n$. Also let $\mathbf{B}(x, r)$ denote the hamming ball of radius r centered at x; that is, $\mathbf{B}(x, r) := \{x' \in [q]^n \,:\, d(x, x') \leq r\}$. In this work, we introduce a nonstandard extension of the notion of distance to small lists of vectors as follows: for $\mathcal{L} \subseteq [q]^n$, define $D_{\max}(x, \mathcal{L}) := \max\{d(x, x') \,:\, x' \in \mathcal{L}\}$ and $D_{\mathrm{avg}}(x, \mathcal{L}) := \mathbf{E}_{x' \in \mathcal{L}}[d(x, x')]$.

We formalize the error recovery capability of the code using list decoding.

Definition 1. *Fix $0 < p < 1/2$ and a positive integer L .*

1. *A q-ary code C is said to be* (p, L)-list decodable *if for all $x \in [q]^n$, we have $|C \cap \mathbf{B}(x, pn)| \leq L - 1$. In other words, for any x and any list $\mathcal{L} \subseteq C$ of size at least L, we have $D_{\max}(x, \mathcal{L}) > pn$.*

2. *C is said to be* (p, L)-strongly list decodable *if for any x and $\mathcal{L}$ as in the previous item, we have $D_{\mathrm{avg}}(x, \mathcal{L}) > pn$.*

3. *C is said to be $(\lambda; p, L)$-(strongly) list decodable if C is (p, L)-(strongly) list decodable, and every codeword in C has weight exactly λn.*

Here the first definition is standard, and the third (i.e., $(\lambda; p, L)$-list decodability) provides a useful notation. Also we emphasize that while formally introduced by us, the notion of (p, L)-

[4] log denotes logarithm to base 2.

strong list-decodability property is implicit in [1, 2, 10]. The following claim asserts that this is a syntactically stronger notion than standard list-decodability:

Proposition 2. *If C is (p, L)-strongly list decodable, then C is (p, L)-list decodable.*

Proof: Follows from the fact that $D_{\max}(x, \mathcal{L})$ always dominates $D_{\text{avg}}(x, \mathcal{L})$ for all x and size-L lists $\mathcal{L}$ of C. ■

Following (and extending) the notation in [8], we make the following definitions to quantify the trade-offs in the different parameters (error-correction radius p, list-size L, weight of the code λ and its rate R). Fix $0 < p, \lambda < 1/2$, $0 \leq R \leq 1$ and a positive integer L. Say that the triple $(p, L; R)$ is *achievable* for (strongly) list decodable codes if there exists (p, L)-(strongly) list decodable codes of rate R for infinitely many lengths n. Similarly the 4-tuple $(\lambda; p, L; R)$ is achievable if there exists $(\lambda; p, L)$-(strongly) list decodable codes of rate R.

Definition 3. *Fix $0 < p < 1/2$.*

1. *Define $L_{p,\gamma}$ to be the least integer L such that $(p, L; 1 - h(p) - \gamma)$ is achievable. Similarly, define $R_{p,L}$ to be the supremum over R such that $(p, L; R)$ is achievable for list decodable codes. Finally, the gap to the optimal/limiting rate (of $1 - h(p)$) is defined to be $\gamma_{p,L} := 1 - h(p) - R_{p,L}$.*
2. *For $\lambda \in (p, 1/2]$, define $R_{p,L}(\lambda)$ to be the supremum rate R for which the 4-tuple $(\lambda; p, L; R)$ is achievable.*

We can also define analogous quantities for strong list decoding, but to prevent notational clutter, we will not explicitly do so.

Useful properties of standard functions. We collect together several facts and estimates that will be useful in our results. The proofs of the standard claims in this subsection will be omitted.

We use the notation $f(n, a, b, i)$ to denote $\frac{\binom{a}{i}\binom{n-a}{b-i}}{\binom{n}{b}}$. We say that a random variable X follows the hypergeometric distribution with parameters n, a, b if $\mathbf{Pr}[X = i] = f(n, a, b, i)$. We will need the following elementary combinatorial identity involving the hypergeometric distribution.

Fact 4. *For all n, a, b, i, we have $f(n, a, b, i) = f(n, b, a, i)$.*

We will use the following estimates related to the binary entropy function without further mention.

Fact 5 (The binary entropy function)**.** *Define the binary entropy function by $h(z) := -z \log z - (1 - z) \log(1 - z)$. Then for any constant $z \in (0, 1)$ and $n \to \infty$, we have $2^{h(z)n - o(n)} \leq \binom{n}{zn} \leq 2^{h(z)n}$.*

Fact 6. *For all $z \in (0, 1)$, we have $z \log(1/z) + (\log e)(z - z^2) \leq h(z) \leq z \log(1/z) + (\log e)z$.*

3 Bounds for strong list decodability

In this section, we establish upper and lower bounds of $1 - h(p) - 1/L^{\Theta(1)}$ on the rate for (p, L)-strongly list decodable codes.

3.1 Lower bound on rate.

The result below follows by a standard random coding argument.

Theorem 7. *Let $0 < p < 1/2$ and L a positive integer. Then for all $\epsilon > 0$ and all sufficiently large lengths n, there exists a (p, L)-strongly list decodable code of rate at least $1 - h(p) - 1/L - \epsilon$.*

Proof: We show that a random code $C : \{0,1\}^{Rn} \to \{0,1\}^n$ of rate $R = 1 - h(p) - 1/L - \epsilon$ is (p, L)-strongly list-decodable whp. For each $m \in \{0,1\}^{Rn}$, pick $C(m)$ independently and uniformly at random from $\{0,1\}^n$. For any $x \in \{0,1\}^n$ and any distinct L-tuple $\{m_1, \ldots, m_L\} \subseteq \{0,1\}^{Rn}$, we are interested in bounding the probability of the event that $D \leq Lpn$, where $D := \sum_{i=1}^{L} d(x, C(m_i))$. Let X be the $\{0,1\}$-string of length Ln obtained by concatenating x repeatedly L times. Similarly, let Y be the $\{0,1\}$-string obtained by concatenating $C(m_1), \ldots, C(m_L)$; then, Y is distributed uniformly at random in $\{0,1\}^{Ln}$ independent of the choice of x. Now, note that D is simply the Hamming distance between X and Y. Hence, the probability that $D \leq pLn$ is at most $2^{(h(p)-1)Ln}$.

Finally, by a union bound over the choice of x and $\{m_1, \ldots, m_L\}$, the probability that the code is not (p, L)-strongly list decodable is at most

$$2^n \binom{2^{Rn}}{L} \cdot 2^{(h(p)-1)Ln} \leq 2^{Ln(\frac{1}{L}+R+h(p)-1)} = 2^{-\epsilon Ln},$$

for the given choice of R, thus establishing the claim. ■

3.2 Upper bound on rate.

We now show an upper bound of $1-h(p)-c_p/L^2$ on the rate of a (p, L)-strongly list decodable code. The proof is based on a simple idea, but to convert this to a full proof requires some calculations and analytic manipulations (involving the hypergeometric distribution and the entropy function). To repeat our main idea from the Introduction, instead of picking the "bad list decoding center" uniformly at random, we pick it randomly very close to a codeword, and this still gives similar guarantees on the number of near-by codewords. Now since the quantity of interest is the average radius, this close-by codeword gives enough savings for us.

Before we proceed with the proof, we first establish a rate upper bound for the special case when all codewords are restricted to be of a fixed weight λn for a suitably chosen $\lambda \in (p, 1/2)$. We can then map this bound to the general case by the following standard argument. (We will establish a converse to this claim in Section 5.)

Lemma 8. *Let $\lambda \in (p, 1/2)$ be such that λn is an integer. If C is a (p, L)-(strongly) list-decodable code of rate $R = 1 - h(p) - \gamma$, then there exists a $(\lambda; p, L)$-(strongly) list decodable code C' of rate at least $h(\lambda) - h(p) - \gamma - o(1)$.*

Proof: For a random center x, the expected number of codewords $c \in C$ with $d(x, c) = \lambda n$ is exactly $|C| \cdot \binom{n}{\lambda n} \cdot 2^{-n} \geq 2^{Rn} \cdot 2^{(h(\lambda)-1-o(1))n} = 2^{(h(\lambda)-h(p)-\gamma-o(1))n}$. Then there exists an x such that the subcode C_x consisting of all codewords at a distance λn from x has a rate at least $h(\lambda) - h(p) - \gamma - o(1)$. Defining C' to be $C_x - x$ gives the claim. ■

We now state our main result establishing a rate upper bound for (p, L)-list decodable codes.

Theorem 9 (Main theorem)**.** *Let $0 < p < 1/2$ and let L a sufficiently large positive integer. Then, there exist $a_p, c_p > 0$ such that the following holds (for sufficiently large lengths n):*

1. *If C is a (p, L)-strongly list-decodable code, then C has rate at most $1 - h(p) - c_p/L^2$.*
2. *For $\lambda := p + a_p/L$, if C is a $(\lambda; p, L)$-strongly list-decodable code, then C has rate at most $h(\lambda) - h(p) - c_p/L^2$.*

Using Lemma 8, it suffices to show the second part. Before we do this, we will establish the following folklore result, whose proof illustrates our idea in a simple case.

Lemma 10 (A warm-up lemma)**.** *If C is a $(\lambda; p, L)$-list-decodable code, then C has rate at most $h(\lambda) - h(p) + o(1)$.*

Proof: The proof is via the probabilistic method. Pick a random subset $S \subseteq [n]$ of coordinates of size αn, with $\alpha := (\lambda - p)/(1 - 2p)$.[5] Define the center x to be the indicator vector of S: $x_i = 1 \iff i \in S$. Let $\mathcal{L}$ be the set of codewords $c \in C$ such that $\mathrm{wt}(c|_S) \geq (1-p)\alpha n$. For any $c \in \mathcal{L}$, we have

$$d(x, c) = (\alpha n - \mathrm{wt}(c|_S)) + \mathrm{wt}(c|_{\overline{S}}) \leq \alpha p n + (\lambda - \alpha(1-p))n = (\lambda - \alpha(1-2p))n,$$

which equals pn for the given choice of α. Hence $\mathcal{L}$ lies entirely inside the ball $\mathbf{B}(x, pn)$.

Now, we want to compute $\mathbf{E}[|\mathcal{L}|]$. For any fixed $c \in C$, the probability that c lies in $\mathcal{L}$ is at least $f(n, \lambda n, \alpha n, \alpha(1-p)n)$, which by Fact 4 equals $\frac{\binom{\alpha n}{(1-p)\alpha n}\binom{(1-\alpha)n}{(\lambda-\alpha(1-p))n}}{\binom{n}{\lambda n}}$. Verify that for our choice of α, it holds that $\lambda - (1-p)\alpha = p(1-\alpha)$. Therefore, conveniently, the above expression is equal to

$$\frac{\binom{\alpha n}{p\alpha n}\binom{(1-\alpha)n}{p(1-\alpha)n}}{\binom{n}{\lambda n}} = \frac{2^{\alpha n h(p) + (1-\alpha)n h(p) - o(n)}}{2^{h(\lambda)n}} = 2^{(h(p) - h(\lambda) - o(1))n}.$$

Therefore, by linearity of expectations, the expected size of $\mathcal{L}$ is at least $|C| \times 2^{(h(p)-h(\lambda)-o(1))n} = 2^{(R + h(p) - h(\lambda) - o(1))n}$. On the other hand, the (p, L)-list decodability of C implies that $|\mathcal{L}| < L$ with probability 1. Comparing the lower and upper bounds on expected size of $\mathcal{L}$, we get $R + h(p) - h(\lambda) - o(1) \leq \frac{1}{n}\log L$, which yields the claim. ■

Proof of Theorem 9: At a high level, we proceed as in the proof of Lemma 10, but in addition to the bad list $\mathcal{L}$, we will produce a special codeword $c^* \in C$ such that $d(x, c^*)$ is much smaller than pn. Then defining a new bad list $\mathcal{L}'$ consisting of c^* and $(L-1)$ other codewords from $\mathcal{L}$, we show that $D_{avg}(x, \mathcal{L}')$ is at most pn, which would contradict the strong list decodability of C.

We now provide the details. Pick a uniformly random codeword $c^* \in C$ and let S be a random subset of $\mathrm{supp}(c^*)$ of size βn, where β is a constant to be chosen appropriately later. Let x be the indicator vector of S. Define $\mathcal{L}$ to be the collection of codewords $c \in C$ such that $\mathrm{wt}(c|S) \geq (1-p)|S|$. (Note that $c^* \in \mathcal{L}$.) Conditioned on c^*, the probability that $c \in \mathcal{L}$ is

$$Q(c^*, c) := \frac{1}{\binom{\lambda n}{\beta n}} \sum_{i=(1-p)\beta n}^{\beta n} \binom{(\lambda-\delta)n}{i}\binom{\delta n}{\beta n - i}$$

where $d(c^*, c) := 2\delta(c^*, c)n = 2\delta n$. Observe that $Q(c^*, c)$ is really a function of $\delta(c^*, c) = \delta$. Therefore, the expected size of $\mathcal{L}$ is $\mathbf{E}_{c^* \in C}\left[\sum_{c \in C} Q(\delta(c^*, c))\right] = |C| \cdot \mathbf{E}_{c, c^* \in C}\left[Q(\delta)\right]$. The following claim lower bounds the expectation of the random variable $Q = Q(\delta)$.

Claim 11 (Estimate of $\mathbf{E}Q$)**.** *There exist $A := (1-p)\log\left(\frac{1-p}{\lambda}\right) + p\log\left(\frac{p}{1-\lambda}\right)$ and $B = B_p \in (0, \infty)$ such that for any code C with all codewords of weight λ, we have*

$$\mathbf{E}_{c^*, c}[Q(\delta(c^*, c))] \geq 2^{-(A\beta + B\beta^2)n}.$$

[5] The reason for setting α to this value will be clear shortly.

Proof Sketch: First, note that $0 \le \delta \le \lambda$ always. Also, it is easy to see that the quantity $Q(\delta)$ is monotonically decreasing with increasing δ. Moreover, by a simple application of the Cauchy-Schwarz inequality, we have $\mathbf{E}_{c^*,c}[\delta] \le \lambda(1-\lambda)$. Now, *if* Q were a *convex* function of δ, then we could lower bound $\mathbf{E}[Q(\delta)]$ by Jensen's inequality; unfortunately, the convexity assumption does not hold. However, it turns out that when δ is restricted to the "middle" range

$$\lambda p + n^{-1/4} \le \delta \le \lambda - \lambda^2/2,$$

we can approximate $Q(\delta)$ well by a *convex* function $\tilde{Q}(\delta)$. Hence the proof strategy can be made to work for $\tilde{Q}$, except for the extreme values of δ. We then handle the "small" regime (i.e., $0 \le \delta \le \lambda p + n^{-1/4}$) and the "large" regime (i.e., $\lambda - \lambda^2/2 \le \delta \le \lambda$) by additional simple tricks.

The complete proof is quite cumbersome since it involves heavy use of several standard estimates (of binomial coefficients) and Taylor approximations. Moreover, one also needs to verify the convexity of $\tilde{Q}(\delta)$. For ease of readability, we finish the rather technical proof in Appendix A. ■

Let us now proceed with completing the proof of Theorem 9. By Claim 11, assuming $R \ge A\beta + B\beta^2 + o(1)$ for a suitable $o(1)$ term, $\mathbf{E}[|\mathcal{L}|] \ge L$. Fix c^* and S such that $|\mathcal{L}| \ge L$. Let $\mathcal{L}'$ be any list containing c^* and $L-1$ other codewords from $\mathcal{L}$. For $c \in \mathcal{L}' \subseteq \mathcal{L}$, we have $d(x,c) \le \beta pn + (\lambda - \beta(1-p))n = (\lambda - \beta(1-2p))n$, whereas $d(x,c^*) = (\lambda-\beta)n$. Averaging these L distances, $D_{\text{avg}}(x,\mathcal{L}') \le (\lambda - \beta(1-2p+2p/L))n$. Now, pick β so that this is at most pn; that is, set $\beta := (\lambda-p)/(1-2p+2p/L)$. For this choice of β, the list $\mathcal{L}'$ contradicts the (p,L)-*strong* list decodability of C. Thus, contrary to our starting assumption, the rate is at most $A\beta + B\beta^2 + o(1)$ (for the special choice of β). We can further upper bound this by (see Claim 23 in Appendix A)

$$h(\lambda) - h(p) - \frac{A_0(\lambda-p)}{L} + B_0(\lambda-p)^2$$

for some $A_0 > 0$ and $B_0 < \infty$ depending on p. Setting $\lambda := p + A_0/(2B_0L)$ gives the claim. ■

4 Bounds for (standard) list decodability

In this section, we consider the rate vs. list size trade-off for the traditional list-decodability notion. For the special case when the fraction of errors is close to $1/2$, [10] showed that any code family of growing size correcting up to $1/2-\gamma$ fraction of errors must have a list size $\Omega(1/\gamma^2)$, which is optimal up to constant factors. When p is bounded away from $1/2$, Blinovsky [1, 3] gives the best known bounds on the rate of a (p,L)-list decodable code. He showed that any code of rate $1-h(p)-\gamma$ has list-size at least $\Omega_p(\log(1/\gamma))$.[6] For completeness we give a self-contained and simpler proof of this result in this section.

Theorem 12 (Blinovsky [1, 3]). *1. Suppose C is $(\lambda; p, L)$-list decodable code with $\lambda = p + \frac{1}{2}p^L$. Then $|C|$ is at most $2L^2/\lambda$ (independent of the blocklength n). (In particular, the rate approaches 0 as $n \to \infty$.)*

2. Suppose C is a (p,L)-list decodable code. Then there exists a constant $c_p > 0$ such that the rate of C is at most $1 - h(p) - 2^{-c_p L}$.

Proof: By Proposition 8, it suffices to show the first part, since then the rate of C is upper bounded by $1 - h(\lambda) = 1 - h(p + \frac{1}{2}p^L) \le 1 - h(p) - \Theta_p(\frac{h'(p)}{2}p^L)$ (using Taylor expansion). We prove the

[6] He states his results in a different form however. The reader is referred to [13] for this form of the result.

first part by the first moment method. Assume that $|C| > 2L^2/\lambda$. Pick a random (distinct) L-tuple of codewords $\mathcal{L} = \{c^1, c^2, \ldots, c^L\} \subseteq C$, and define x by $x_i = 1$ iff $c^j_i = 1$ for all $1 \le j \le L$. Note that x is at a distance of $\lambda n - \mathrm{wt}(x)$ from each c^j, so that $\mathbf{E}[D_{\max}(x, \mathcal{L})] = \lambda n - \mathbf{E}[\mathrm{wt}(x)]$. Thus to complete the proof, it suffices to show that $\mathbf{E}[\mathrm{wt}(x)] \ge \frac{1}{2}p^L$.

Define the function $\vartheta : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ by $\vartheta(z) = \binom{\max\{z, L-1\}}{L}$. By standard closure properties of convex functions, ϑ is convex on $\mathbb{R}_{\ge 0}$. Now, let $M := |C|$ and M_i be the number of codewords with 1 in the i^{th} position. Then it can be verified that $x_i = 1$ with probability $\vartheta(M_i)/\binom{M}{L}$. Thus, by linearity of expectations, the expected weight of x is

$$\frac{1}{\binom{M}{L}} \sum_{i=1}^{n} \vartheta(M_i) = \frac{n}{\binom{M}{L}} \mathbf{E}_{i \in [n]}\left[\vartheta(M_i)\right] \overset{(a)}{\ge} \frac{n}{\binom{M}{L}} \vartheta(\mathbf{E}[M_i]) \overset{(b)}{=} \frac{n}{\binom{M}{L}} \binom{\lambda M}{L}.$$

Here we have used (a) Jensen's inequality, and (b) the fact that $\lambda M > 2L^2 \ge L$. Finally, a straightforward approximation gives the promised bound:

$$\frac{\binom{\lambda M}{L}}{\binom{M}{L}} \ge \frac{(\lambda M - L)^L}{M^L} = \lambda^L \left(1 - \frac{L}{\lambda M}\right)^L \ge \lambda^L \left(1 - \frac{L^2}{\lambda M}\right) \ge \frac{1}{2}\lambda^L \ge \frac{1}{2}p^L .$$

■

The above method can be adapted for q-ary codes with an additional trick.

Theorem 13. *1. Suppose C is a q-ary $(\lambda; p, L)$-list decodable code with $\lambda = p + \frac{1}{2L}p^L$. Then $|C|$ is at most $2L^2/\lambda$.*

2. Suppose C is a q-ary (p, L)-list decodable code. Then there exists a constant $c_{p,q} > 0$ such that the rate of C is at most $1 - h_q(p) - 2^{-c_{p,q}L}$.

Before we prove Theorem 13, we will state a convenient lemma due to Erdös. (See Section 2.1 of [12] for reference.) This result was implicitly established in our proof of Theorem 12; so we will omit the formal proof.

Lemma 14 (Erdös 1964)**.** *Suppose $\mathcal{A}$ is a set system over the ground set $[n]$, such that each $A \in \mathcal{A}$ has size at least λn. Then if $|\mathcal{A}| \ge 2L^2/\lambda$, then there exist distinct $A_1, A_2, \ldots, A_L$ in $\mathcal{A}$ such that $\bigcap_{i=1}^{L} A_i$ has size at least $\frac{1}{2}n\lambda^L$.*

Proof of Theorem 13: As in Theorem 12, it suffices to show the first part. Towards a contradiction, assume $|C| \ge 2L^2/\lambda$. Define the set system $\mathcal{A} = \{\mathrm{supp}(c) : c \in C\}$. By Lemma 14, there exists an L-tuple $\{c^1, c^2, \ldots, c^L\}$ of codewords such that the intersection of their support, say S, has size $\ge \frac{1}{2}n\lambda^L \ge \frac{1}{2}np^L$. Arbitrarily partition the coordinates in S into L parts $\{S_1, \ldots, S_L\}$ of almost-equal size $n/(2L) \cdot p^L$. Now, define the center x by:

$$x_i := \begin{cases} c^j_i, \text{ if } i \in S_j, \text{ and} \\ 0, \text{ if } i \notin S. \end{cases}$$

Note that x agrees with c^j on S_j, so that $d(x, c^j) \le \lambda n - \frac{1}{2L}p^L n = pn$. Therefore, $\{c^1, \ldots, c^L\}$ is a bad list of codewords contradicting the (p, L)-list decodability of C. ■

5 Constant-weight vs. General codes

In this section, we will understand the rate vs. list-size trade-offs for constant-weight codes, that is, codes with every codeword of weight λn, where $\lambda \in (p, 1/2]$ is a parameter. (Note that setting

$\lambda = 1/2$ corresponds to arbitrary codes having no weight restrictions.) As observed earlier, a typical approach in coding theory to establish rate upper bounds is to study the problem under the above constant-weight restriction. One then proceeds to show a strong negative result of the flavor that a code with the stated properties must have a constant size (and in particular *zero* rate). For instance, the first part of Theorem 12 above is of this form. Finally, mapping this bound to arbitrary codes, one obtains a rate upper bound of $1 - h(\lambda)$ for the original problem. (Note that Lemma 8 provides a particular formal example of the last step.)

In particular, Blinovsky's rate upper bound (Theorem 12)[7] of $1 - h(p) - 2^{-O(L)}$ for (p, L)-list decodable codes follows this approach. More precisely, he proves that, under the weight-λ restriction, such code must have zero rate for all $\lambda \leq p + 2^{-cL}$ for some $c < \infty$. One may then imagine improving the rate upper bound to $1-h(p)-L^{-O(1)}$ *simply by* establishing the latter result for correspondingly higher values of λ (i.e., up to $p + L^{-O(1)}$). We show that this approach cannot work by establishing that list-decodable codes of positive (but possibly small) rates exist as long as $\lambda - p \geq 2^{-O(L)}$. Thus Blinovsky's result identifies the correct *zero-rate regime* for the list-decoding problem; in particular, his bound is also the best possible if we restrict ourselves to this approach.

In the opposite direction, we show that the task of establishing rate upper bounds for constant weight codes is not significantly harder than the general problem. Formally, we state that that if the "gap to capacity" for general codes is γ, then the gap to capacity for weight-λ codes is *at least* $\gamma\left(\frac{\lambda-p}{1/2-p}\right)$. Stated differently, if our goal is to establish a $L^{-O(1)}$ lower bound on the gap γ, then we do not lose by first passing to a suitable λ (that is not too close to p).

5.1 Zero-rate regime

We now prove the existence of (p, L)-strongly list-decodable codes of positive rate where all codewords have constant weight which is very close to pn.

Theorem 15. *For every $0 < p < 1/2$, there exists $d = d(p) = \frac{1}{2}(1/2 - p)^2 \in (0, \infty)$ such that for all sufficiently large L, there exists a $(\lambda; p, L)$-strongly list decodable code of rate at least $R - o(1)$ with $R = e^{-2dL}$ and $\lambda \in [p, p + 12e^{-dL}]$.*

The proof proceeds by random coding followed by expurgation. Set $\epsilon := 4e^{-dL}$ and $\lambda' := p+2\epsilon$. Now, pick a random $2^{Rn} \times n$ *code matrix* C with each entry set to 1 with probability λ'. For our choice of parameters, we can show that whp, C satisfies the following properties:

- C is (p, L)-strongly list-decodable.
- Every codeword has weight $(\lambda' \pm \epsilon)n$. In particular, the maximum weight is at most $(p+3\epsilon)n$.

Pick a C satisfying these two properties, and let C_i denote the sub-code of C consisting of the weight-i codewords. Then, defining $i^* = \lambda n$ to be the most popular weight, the subcode C_{i^*} satisfies our constraints. The formal proof follows.

Proof of Theorem 15: Set $\epsilon := 4e^{-dL}$ and $\lambda' := p + 2\epsilon$. Assume that L is large enough so that $p + 4\epsilon < 1/2$ and verify that $1/2 - \lambda' \geq \frac{1}{2}(1/2 - p)$ in this case. Pick a random code $C : \{0,1\}^{Rn} \to \{0,1\}^n$, where for each $y \in \{0,1\}^{Rn}$, every coordinate of $C(y)$ is chosen independently to be 1 with probability λ'. First, by Chernoff bound followed by union bound, the probability that there exists $y \in \{0,1\}^{Rn}$ with $|\mathrm{wt}(y) - \lambda' n| > \epsilon n$ is at most $2^{Rn} \cdot 2^{-2\epsilon^2 n}$. This is our first bad event.

[7] For notational ease, we supress the dependence on p in the O and Ω notations in this informal discussion.

Now, we bound the probability of the occurrence of a bad list of codewords. Fix a list $\{y_1, \ldots, y_L\} \subseteq (\{0,1\}^{Rn})^L$ and define x to be its centroid: that is, x_j is the majority of the L bits $(C(y_i)_j : 1 \leq i \leq L)$. By Chernoff bound, for $j \in [n]$, the probability that $x_j = 1$ is at most

$$e^{-2(1/2-\lambda')^2 L} \leq e^{-\frac{1}{2}(1/2-p)^2 L} = e^{-Ld} = \epsilon/4.$$

By a second application of Chernoff bound, the probability that the weight of x exceeds $\epsilon n = (1+3)(\epsilon n/4)$ is at most $e^{-\frac{3^2(\epsilon n/4)}{3}} = e^{-3\epsilon n/4}$. Our second bad event is that there exists a list $\{y_1, \ldots, y_L\}$ such that the weight of x is $> \epsilon n$. By union bound over all possible lists, the probability of this event is at most $\binom{2^{Rn}}{L} \cdot e^{-3\epsilon n/4} \leq e^{(RL-3\epsilon/4)n}$.

Since $R < \min\{\epsilon^2, \epsilon/2L\}$, the random code avoids both the bad events with probability $1 - 2^{-\Omega_p(n)}$. Fix any such code C (avoiding both bad events). For any list $\{y_1, \ldots, y_L\} \subseteq C$ with centroid x, for all $1 \leq i \leq L$, we have

$$d(x, y_i) \geq \mathrm{wt}(y_i) - \mathrm{wt}(x) \geq (\lambda' - \epsilon)n - \epsilon n = (\lambda' - 2\epsilon)n = pn,$$

where x is the center of the list as defined above. Therefore, the average distance of the list from the center is also at least pn. Hence, the code C' is (p, L)-strongly list decodable. Now, using the pigeonhole principle, we can find a sub-code with all codewords having weight exactly w having size at least $2^{Rn}/(n+1) = 2^{(R-o(1))n}$. Defining $\lambda := w/n$, we obtain a $(\lambda; p, L)$-strongly list decodable code of rate $R - o(1)$. Finally, it is clear that $\lambda \leq \lambda' + \epsilon = p + 3\epsilon \leq p + 12e^{-Ld}$. ■

5.2 A reverse connection between constant-weight and arbitrary codes

Lemma 16. *Let $\gamma = \gamma_{p,L}$ be the gap to capacity for arbitrary codes. Then, for every $\lambda \in (p, 1/2]$,*

$$h(\lambda) - h(p) - \gamma \leq R_{p,L}(\lambda) \leq h(\lambda) - h(p) - \gamma\left(\frac{\lambda - p}{1/2 - p}\right).$$

Proof: The left inequality is essentially the content of Claim 8; we show the second inequality here. Suppose C is a $(\lambda; p, L)$-list decodable code of rate R. Pick a random subset S of coordinates of size αn with $\alpha = (\lambda - p)/(1/2 - p)$. (The motivation for this choice will become clear shortly.) Consider the subcode C' consisting of the codewords $c \in C$ such that $\mathrm{wt}(c|_S) \geq \alpha n/2$. For our choice of α, one can verify that if $c \in C'$, then c has weight at most $p(1-\alpha)n = p|\overline{S}|$ when restricted to $\overline{S}$.

The key insight is that the code $C'_{|S} := \{c|_S : c \in C'\}$ (of blocklength αn) is (p, L)-list decodable. Suppose not. Then there exists a center $x' \in \{0,1\}^S$ and a size-L list $\mathcal{L} \subseteq C'$ such that $d(x', c|_S) \leq p\alpha n$ for all $c \in \mathcal{L}$. Now, extend x' to $x \in \{0,1\}^n$ such that $x|_S = x'$ and x_i is zero for $i \notin S$. Then, for $c \in \mathcal{L}$, we have $d(x, c) \leq d(x', c|_S) + \mathrm{wt}(c|_{\overline{S}}) \leq p\alpha n + p(1-\alpha)n = pn$. Thus, $\mathcal{L} \subseteq \mathbf{B}(x, pn)$, contradicting the (p, L)-list decodability of C (and hence of C').

By hypothesis, we can bound the size of $C'_{|S}$ by $2^{(1-h(p)-\gamma)\alpha n}$ (with probability 1). On the other hand, in expectation, the size of $C'_{|S}$ is at least

$$|C| \cdot \frac{\binom{\lambda n}{\alpha n/2}\binom{(1-\lambda)n}{\alpha n/2}}{\binom{n}{\alpha n}} = |C| \cdot \frac{\binom{\alpha n}{\alpha n/2}\binom{(1-\alpha)n}{(\lambda-\alpha/2)n}}{\binom{n}{\lambda n}}$$

appealing to Fact 4 again. Finally, verify that $\lambda - \alpha/2 = p(1-\alpha)$. By standard approximation, this quantity is at least $\mathsf{exp2}[R + \alpha + (1-\alpha)h(p) - h(\lambda) - o(1))n]$.[8] Comparing the upper and

[8] We use $\mathsf{exp2}(z)$ to denote 2^z.

lower bound on the (expected) size of $C'_{|S}$, we get $R + \alpha + (1-\alpha)h(p) - h(\lambda) \leq (1 - h(p) - \gamma)\alpha$. Rearranging this inequality gives the desired bound $R \leq h(\lambda) - h(p) - \alpha\gamma$. ■

6 List-size Bounds for Random codes

In this section, we establish optimal (up to constant factors) bounds on the list-size of random codes, both general as well as linear. Results of this vein were already shown by Rudra for the errors case [13], based on the large near-disjoint packings of Hamming balls implied by Shannon's capacity theorems. Here we give a direct proof based on the second moment method.[9] In addition, our proofs extend easily to give list-size bounds for the erasures case as well.

By a random code, we mean a random map $\mathbf{Enc} : \{0,1\}^k \to \{0,1\}^n$ where the image $\mathbf{Enc}(x)$ of each $x \in \{0,1\}^k$ is picked independently and uniformly at random from $\{0,1\}^n$. On the other hand, to obtain a random *linear* code, we fix an arbitrary basis for the vector space $\{0,1\}^k$, and the encoding of the basis vectors is chosen independently and uniformly at random. The encoding map $\mathbf{Enc}$ is then extended for all messages in $\{0,1\}^k$ via linearity.

6.1 Bounds for Random codes under Errors

As mentioned before, our results proceed directly via the second moment method. Towards this goal, we define a random variable $\mathcal{Z}$ that counts the number of *witnesses* (i.e., a bad list of codewords together with the center) that certify the violation of the (p, L)-list decodability property. Note that the code is (p, L)-list decodable iff $\mathcal{Z} = 0$. We then show that $\mathcal{Z}$ has large expectation (i.e., exponential in n) and that $\mathbf{Var}[\mathcal{Z}] = \exp(-\Omega_{q,p,\gamma}(n))\mathbf{E}[\mathcal{Z}]^2 = o(\mathbf{E}[\mathcal{Z}]^2)$. Using the Chebyshev inequality, we can conclude that $\mathcal{Z} > 0$, except with an exponentially small probability, which gives the claim.

As a particular example, consider the case of random general codes under errors. Here, we let X be an arbitrary distinct L-tuple of messages $\{x_1, x_2, \ldots, x_L\} \subseteq C$ and a be an arbitrary center. Then define the indicator random variable $I(X, a)$ for the event that $d(a, \mathbf{Enc}(x)) \leq pn$ for all $x \in X$. Finally, define $\mathcal{Z} := \sum_{X,a} I(X, a)$. The mean and variance estimates of $\mathcal{Z}$ follow by a standard calculation. Our formal results and proofs follow.

6.1.1 General codes

Theorem 17. *For every $0 < p < 1 - 1/q$ and $\gamma > 0$, with probability $1 - q^{-\Omega_{p,\gamma}(n)}$, a random q-ary code of rate $1 - h_q(p) - \gamma$ is not $(p, \frac{1-h_q(p)}{2\gamma})$-list decodable.*

Before presenting the proof, let us define some convenient notation. We denote by $\mathbf{B}_q(a, r)$ the Hamming ball with center a and radius r. Define $\mathrm{Vol}_q(n, r)$ be the volume of $\mathbf{B}_q(\cdot, r)$, and $\mu_q(n, r) := \mathrm{Vol}(n, p)/q^n$. It is a standard fact that $q^{(H_q(z)-o(1))n} \leq \mathrm{Vol}_q(n, zn) \leq q^{H_q(z)n}$. We will use $\mathbf{B}(a)$ (resp. μ) as a shorthand to denote $\mathbf{B}_q(a, pn)$ (resp. $\mu_q(n, pn)$).

Proof: At a high level, we apply the second moment method to the random variable $\mathcal{Z}$ that counts the number of *witnesses* (i.e., a bad list of codewords and the corresponding center) certifying the violation of the (p, L)-list decodability property. Consider a random code $\mathbf{Enc} : [q]^{k=Rn} \to [q]^n$

[9] We remark that the argument in [13] is also based on the second moment method, but applied to a more complicated random variable.

with $R = 1 - h_q(p) - \gamma$. For a list of L messages $X = \{x_1, x_2, \ldots, x_L\} \subseteq \{0,1\}^k$, and $a \in \{0,1\}^n$, define the indicator variable $I(X, a)$ to be 1 iff $\mathbf{Enc}(x) \in \mathbf{B}(a)$ for all $x \in X$. Then define $\mathcal{Z} := \sum_{X,a} I(X, a)$. Clearly, $\mathcal{Z} > 0$ iff the code is (p, L)-list decodable.

For every x and every a, the event $\mathbf{Enc}(x) \in \mathbf{B}(a)$ occurs wp $\mu_q(n, pn) = \mu$; therefore, by independence, for a list of messages X, we have $\mathbf{E}[I(X, a)] = \mu^L$. Therefore, by the linearity of expectations, $\mathbf{E}[\mathcal{Z}] = \mu^L \binom{2^k}{L} q^n \geq L^{-L} \left(q^k \mu\right)^L q^n$. For two lists of messages X and Y, say $X \sim Y$ if $X \cap Y \neq \emptyset$. Clearly, if $X \not\sim Y$, then the events $I(X, a)$ and $I(Y, b)$ are independent. Therefore, we have

$$\begin{aligned}
\mathbf{Var}[\mathcal{Z}] &= \sum_{X,Y} \sum_{a,b} (\mathbf{E}[I(X,a)I(Y,b)] - \mathbf{E}[I(X,a)]\mathbf{E}[I(Y,b)]) \\
&\leq \sum_{X \sim Y} \sum_{a,b} \mathbf{E}[I(X,a)I(Y,b)] = \sum_{X \sim Y} \sum_{a,b} \mathbf{Pr}[I(X,a) = 1 \text{ and } I(Y,b) = 1] \\
&= q^{2n} \sum_{X \sim Y} \mathbf{Pr}_{a,b,\mathbf{Enc}}[I(X,a) = 1 \text{ and } I(Y,b) = 1],
\end{aligned}$$

where, in addition to the randomness in the code, the centers a and b are also picked at random.

Fix a pair (X, Y) such that $|X \cap Y| = \ell > 0$. Let $z \in X \cap Y$ be arbitrary. Then for any a, b, the event $I(X, a) = I(Y, b) = 1$ implies that

- $\mathbf{Enc}(x) \in \mathbf{B}(a)$ for $x \in X \backslash \{z\}$;
- $\mathbf{Enc}(y) \in \mathbf{B}(b)$ for $y \in Y \backslash X$;
- $\{a, b\} \subseteq \mathbf{B}(\mathbf{Enc}(z))$.

Thus this event happens with probability at most $\mu^{2L-\ell+1}$. Finally, summing over all the pairs (X, Y) with $\ell > 0$ (the number of such pairs is at most $L^{2L} q^{k(2L-\ell)}$),

$$\mathbf{Var}[\mathcal{Z}] \leq q^{2n} \sum_{\ell=1}^{L} L^{2L} q^{k(2L-\ell)} \mu^{2L-\ell+1} \leq \sum_{\ell=1}^{L} L^{4L} (q^k \mu)^{-\ell} \mu \cdot (\mathbf{E}\mathcal{Z})^2,$$

after some rearrangement. Note that $q^k \mu = q^{-\gamma n}$ for our choice of the rate. Therefore,

$$\mathbf{Var}[\mathcal{Z}] \leq \sum_{\ell=1}^{L} L^{4L} q^{\gamma \ell n} \mu \cdot (\mathbf{E}\mathcal{Z})^2 \leq L^{4L+1} q^{\gamma L n - (1 - h_q(p))n} \cdot (\mathbf{E}\mathcal{Z})^2.$$

Therefore, letting $L = (1 - h_q(p))/(2\gamma)$, we observe that $\mathbf{Var}[\mathcal{Z}] = q^{-\Omega_{p,\gamma}(n)} (\mathbf{E}\mathcal{Z})^2$. Finally, by Chebyshev's inequality, $\mathcal{Z} = 0$ (i.e., the code is (p, L)-list decodable) with probability $q^{-\Omega_{p,\gamma}(n)}$. ■

6.1.2 Random linear codes

We now turn to the case of random linear codes.

Theorem 18. *For every $0 < p < 1 - 1/q$ there exists $\delta_{q,p} > 0$ such that for all $\gamma > 0$, with probability $1 - q^{-\Omega_{p,\gamma}(n)}$, a random q-ary linear code of rate at least $1 - H_q(p) - \gamma$ is not $(p, \delta_{q,p}/(2\gamma))$-list decodable with high probability.*

Proof: We follow the same outline as in Theorem 17; we will only highlight the differences. Consider a random linear code of dimension $k = (1 - H_q(p) - \gamma)n$. We define $I(X, a)$ in an identical manner, but only for the *linearly independent* lists of messages X. (The definition of $\mathcal{Z}$ is unchanged.) Furthermore, for a pair of lists X and Y, define $\ell = \ell(X, Y) := \dim(\mathrm{span}(X) \cap \mathrm{span}(Y))$ (rather than the size of their intersection). Moreover, we say that $X \sim Y$ iff $\ell = 0$; that is, iff $\mathrm{span}(X)$ and $\mathrm{span}(Y)$ have a nontrivial intersection.

Estimating $\mathbf{E}[\mathcal{Z}]$ as before[10], we get $\mathbf{E}[\mathcal{Z}] \geq \frac{1}{2} \cdot L^{-L}(q^k\mu)^L q^n$. Also, we can write

$$\mathbf{Var}[\mathcal{Z}] \leq q^{2n} \sum_{X \sim Y} \mathbf{Pr}_{a,b,\mathbf{Enc}} \left[I(X, a) = 1 \text{ and } I(Y, b) = 1 \right].$$

Fix a pair X, Y such that $\dim(\mathrm{span}\ X \cap \mathrm{span}\ Y) = \ell > 0$. Then, there exists $Z \subseteq Y$ of size $L - \ell$, such that $X \sim Z$ and $Y \subseteq \mathrm{span}(\ X \cup Z)$. Let $y_0 \in Y \backslash Z$ be arbitrary. Then, since $y_0 \in \mathrm{span}(\ X \cup Z)$, we have $y_0 = \sum_{u \in X \cup Z} \zeta(u) u$ for some scalars $\zeta(u)$. Note that it is possible that y_0 lies in the span of X. But, since Y is an independent set, y_0 cannot be written as a linear combination of vectors from $Z \subseteq Y \backslash \{y_0\}$. Hence, there exists some $u \in X$ with $\zeta(u) \neq 0$.

In order to compute the desired probability, condition on the event that $\mathbf{Enc}(u) \in \mathbf{B}(a)$ for $u \in X$ and $\mathbf{Enc}(u) \in \mathbf{B}(b)$ for $u \in \mathbf{B}(b)$. We may re-express this as $\mathbf{Enc}(u) = \delta(u) + a$ for $u \in X$ and $\mathbf{Enc}(u) = \delta(u) + b$ for $u \in Z$. We thus get a family of iid random variables $\{\delta(u)\}_{u \in X \cup Z}$, each of which is uniformly distributed inside $\mathbf{B}(\mathbf{0})$. Further they are also independent of a and b. In terms of the $\delta(\cdot)$'s, we have $\mathbf{Enc}(y_0) - b = \sum_{u \in X \cup Z} \zeta(u)\delta(u) + \zeta(X)\, a + (\zeta(Z) - 1)\, b$.

We claim that the conditional probability that $\mathbf{Enc}(y_0) - b \in \mathbf{B}(\mathbf{0})$ is at most $q^{-\delta_{p,q} n}$. We discuss two cases:

1. Suppose $\zeta(X) \neq 0$ *or* $\zeta(Z) \neq 1$. Then conditioned on $\delta(\cdot)$'s, the random variable $\mathbf{Enc}(y_0) - b$ is distributed uniformly at random and hence falls inside $\mathbf{B}(\mathbf{0})$ with probability μ.

2. Suppose $\zeta(X) = 0$ and $\zeta(Z) = 1$. In this case, $\mathbf{Enc}(y_0) - b$ is simply a sum of some number of points uniformly sampled from the ball $\mathbf{B}(\mathbf{0})$. Notice that since Y is not linearly dependent, we must have $\zeta(x) \neq 0$ for some $x \in X$. Also, since $\zeta(x)$'s sum to zero, there are at least two nonzero $\zeta(x)$'s. Therefore, $\mathbf{Enc}(y_0) - b$ is the sum of $l \geq 2$ random points chosen uniformly from $\mathbf{B}(\mathbf{0})$. We use the following fact: that there exists $\delta_{q,p} > 0$ such that, if $w_1, w_2, \ldots, w_l$ are $l \geq 2$ independent and uniformly random samples from $\mathbf{B}(\mathbf{0})$, then the probability that $w_1 + w_2 + \ldots + w_l$ is also inside $\mathbf{B}(\mathbf{0})$ is bounded by $q^{-\delta_{q,p} n}$. Thus, the stated event also occurs with probability $q^{-\delta_{q,p} n}$. (Without loss of generality, we may take $q^{-\delta_{q,p} n}$ to be larger than μ.)

Therefore, the conditional probability is at most $2^{-\delta_{q,p} n}$. Thus, $\mathbf{Var}[\mathcal{Z}] \leq q^{2n} \sum_{X \sim Y} \mu^{2(L-\ell)} q^{-\delta_{q,p} n}$. Proceeding as before, we get $\mathbf{Var}[\mathcal{Z}] \leq O(L^{4L+1} q^{(\gamma L - \delta_{q,p})n} \mathbf{E}[\mathcal{Z}]^2)$. The conclusion follows similarly. ■

6.2 Bounds for Random codes under Erasures

To model erasures, we augment the alphabet $[q]$ with the *erasure symbol* $*$ to get $[q]_* := [q] \cup \{*\}$. For $a \in [q]_*^n$, define $\mathrm{supp}^*(a)$ to be the set of all indices i such that $a_i \neq *$. Let $\mathcal{E}_q(n, r)$ be the set of $a \in [q]_*^n$ such that $|\mathrm{supp}^*(a)| = n - r$. Say that $a, b \in [q]_*^n$ agree with each other if $a_i = b_i$ for all $i \in \mathrm{supp}^*(a) \cap \mathrm{supp}^*(b)$.

[10] Here we must be careful to sum over only the linearly independent L-tuples X.

Definition 19. *A code $C \subseteq \{0,1\}^n$ is said to be (p, L)-erasure list decodable if for all $a \in \mathcal{E}_q(n, pn)$, at most $L-1$ codewords in C (treated as strings over $[q]_*$) agree with a.*

We now state our results showing limitations of erasure list-decodability of random and random linear codes.

Theorem 20. *For every $0 < p < 1$ and $\gamma > 0$, with probability $1 - q^{-\Omega_{p,\gamma}(n)}$, a random code of blocklength n and rate at least $1 - p - \gamma$ is not $(p, \frac{1-p}{2\gamma})$-erasure list decodable.*

Theorem 21. *Let q be a prime power. Then there exists a constant $c_q > 0$ such that for every $0 < p < 1$ and $\gamma > 0$, with probability $1 - q^{-\Omega_{p,\gamma}(n)}$, a random q-ary linear code of rate at least $1 - p - \gamma$ is not $(p, q^{\frac{c_q p(1-p)}{2\gamma}})$-erasure list decodable with high probability.*

Note the exponential gap in the list size for linear and general codes under erasures. We present the proofs for the erasure case in Appendix B.

References

[1] V. M. Blinovsky. Bounds for codes in the case of list decoding of finite volume. *Problems of Information Transmission*, 22(1):7–19, 1986. 1, 2, 3, 5, 8

[2] V. M. Blinovsky. *Asymptotic Combinatorial Coding Theory*. Kluwer Academic Publishers, Boston, 1997. 5

[3] V. M. Blinovsky. Code bounds for multiple packings over a nonbinary finite alphabet. *Problems of Information Transmission*, 41(1):23–32, 2005. 1, 8

[4] V. M. Blinovsky. On the convexity of one coding-theory function. *Problems of Information Transmission*, 44(1):34–39, 2008. 1

[5] P. Elias. List decoding for noisy channels. *Technical Report 335, Research Laboratory of Electronics, MIT*, 1957. 1

[6] P. Elias. Error-correcting codes for list decoding. *IEEE Transactions on Information Theory*, 37:5–12, 1991. 1

[7] V. Guruswami. Linear-algebraic list decoding of folded Reed-Solomon codes. In *Proceedings of the 26th IEEE Conference on Computational Complexity*, pages 77–85, June 2011. 1

[8] V. Guruswami, J. Håstad, and S. Kopparty. On the list-decodability of random linear codes. *IEEE Transactions on Information Theory*, 57(2):718–725, 2011. 1, 5

[9] V. Guruswami and A. Rudra. Explicit codes achieving list decoding capacity: Error-correction up to the Singleton bound. *IEEE Transactions on Information Theory*, 54(1):135–150, January 2008. 1

[10] V. Guruswami and S. P. Vadhan. A lower bound on list size for list decoding. *IEEE Transactions on Information Theory*, 56(11):5681–5688, 2010. 2, 3, 5, 8

[11] V. Guruswami and C. Wang. Optimal rate list decoding via derivative codes. In *Proceedings of APPROX/RANDOM 2011*, pages 593–604, August 2011. 1

[12] S. Jukna. *Extremal Combinatorics: with applications in Computer Science*. Springer, 2001. 9

[13] A. Rudra. Limits to list decoding of random codes. *IEEE Transactions on Information Theory*, 57(3):1398–1408, 2011. 2, 8, 12

[14] J. M. Wozencraft. List Decoding. *Quarterly Progress Report, Research Laboratory of Electronics, MIT*, 48:90–95, 1958. 1

[15] V. V. Zyablov and M. S. Pinsker. List cascade decoding. *Problems of Information Transmission*, 17(4):29–34, 1981 (in Russian); pp. 236-240 (in English), 1982. 1, 2

A Rate upper bound for strong list decoding

We now finish the proof of Claim 11 which was used in the proof of our main result (Theorem 9) on the rate upper bound for strongly list-decodable codes.

Proof of Claim 11: (Continued from Section 3.2)
Divide the range of δ into three regimes.

Small δ: $0 \le \delta \le \lambda p + n^{-1/4}$. We claim that in this regime, $Q(\delta) \ge 2^{-o(n)}$. To see this, set

$$i := \begin{cases} \frac{\beta}{\lambda}(\lambda - \delta)n, & \text{if } 0 \le \delta \le \lambda p, \text{ and} \\ \beta(1-p)n, & \text{if } \lambda p \le \delta \le \lambda p + n^{-1/4}. \end{cases}$$

It is easy to see that $i \ge \beta(1-p)n$ and that $i = \frac{\beta}{\lambda}(\lambda-\delta)n + o(n)$ for all δ. (Here, $\frac{\beta}{\lambda}(\lambda-\delta)n$ represents the expected weight of S.) Now, $Q(\delta)$ is at least

$$\frac{1}{\binom{\lambda n}{\beta n}} \binom{(\lambda-\delta)n}{i} \binom{\delta n}{\beta n - i}.$$

For the prescribed choice of i, by Stirling's approximation, we can verify that $Q(\delta) \ge 2^{-o(n)}$.

Large δ: $\lambda - \lambda^2/2 \le \delta \le \lambda$. In this case, $Q(\delta)$ can be very small, which affects the expectation. However, we can upper bound the probability of this event by Markov inequality:

$$\mathbf{Pr}[\delta \ge \lambda - \lambda^2/2] \le \frac{\lambda - \lambda^2}{\lambda - \lambda^2/2} \le 1 - \lambda/2.$$

Therefore, δ is smaller than $\lambda - \lambda^2/2$ with probability at least $\lambda/2$.

Middle regime: $\lambda p + n^{-1/4} \le \delta \le \lambda - \lambda^2/2$. In this regime, we have

$$Q(\delta) \ge \frac{1}{\binom{\lambda n}{\beta n}} \binom{(\lambda-\delta)n}{(1-p)\beta n} \binom{\delta n}{p\beta n}.$$

Expressing this probability in terms of "rate", we get

$$\frac{1}{n} \log Q(\delta) \ge (\lambda - \delta) h\left(\frac{(1-p)\beta}{\lambda - \delta}\right) + \delta h\left(\frac{p\beta}{\delta}\right) - \lambda h\left(\frac{\beta}{\lambda}\right) - o(1).$$

Lower bounding this using Fact 6, we get

$$\begin{aligned}\frac{1}{n}\log Q(\delta) &\geq \beta[(1-p)\log\frac{\lambda-\delta}{(1-p)} + p\log\frac{\delta}{p} - \log\lambda] - \beta^2(\log e)\left(\frac{(1-p)^2}{\lambda-\delta} + \frac{p^2}{\delta}\right) - o(1).\\ &= \beta[(1-p)\log(\lambda-\delta) + p\log\delta + h(p) - \log\lambda] - \beta^2(\log e)\left(\frac{(1-p)^2}{\lambda-\delta} + \frac{p^2}{\delta}\right).\end{aligned}$$

When δ is restricted to the middle regime, verify that $\frac{(1-p)^2}{\lambda-\delta} + \frac{p^2}{\delta} = O_p(1)$, independent of λ and δ. Therefore,

$$Q(\delta) \geq (\lambda-\delta)^{\beta(1-p)n}\,\delta^{\beta pn}\,2^{\left(\beta h(p) - \beta\log\lambda - O_p(\beta^2)\right)n},$$

which we define to be $\tilde{Q}(\delta)$. Note that, conveniently, $\tilde{Q}(\delta)$ is a *polynomial* function of δ.

The key claim is that in the desired range $\lambda p + n^{-1/4} \leq \delta \leq \lambda - \lambda^2/2$, $\tilde{Q}(\delta)$ is a both *monotonically decreasing* and *convex* in δ. Clearly, it suffices to show these two properties for the function $\tilde{\tilde{Q}}(\delta) := (\lambda-\delta)^{\tau_1}\delta^{\tau_2}$, where we have set $\tau_1 := \beta(1-p)n$ and $\tau_2 := \beta pn$ for ease of notation.

1. **Monotonicity.** Differentiating the function wrt δ, we get

   $$\frac{\mathrm{d}}{\mathrm{d}\delta}\tilde{\tilde{Q}}(\delta) = (\lambda-\delta)^{\tau_1-1}\delta^{\tau_2-1}\,[\tau_2(\lambda-\delta) - \tau_1\delta].$$

   For our parameters, $\tau_2(\lambda-\delta) - \tau_1\delta = \beta n[p\lambda - \delta] \leq -\beta n^{3/4} < 0$. Thus $\tilde{Q}$ is monotonically decreasing.

2. **Convexity.** Differentiating twice wrt δ, we get

   $$\frac{\mathrm{d}^2}{\mathrm{d}\delta^2}\tilde{\tilde{Q}}(\delta) = (\lambda-\delta)^{\tau_1-2}\delta^{\tau_2-2}\,[(\tau_1\delta - \tau_2(\lambda-\delta))^2 - \tau_1\delta^2 - \tau_2(\lambda-\delta)^2].$$

   For our choice of τ_1 and τ_2, this simplifies to

   $$(\lambda-\delta)^{\tau_1-2}\delta^{\tau_2-2}\,[\beta^2n^2(\delta-p\lambda)^2 - \beta n(1-p)\delta^2 - \beta np(\lambda-\delta)^2] \geq (\lambda-\delta)^{\tau_1-2}\delta^{\tau_2-2}\,[\beta^2n^2(\delta-p\lambda)^2 - 2n].$$

   Finally, since $\delta - p\lambda \geq n^{-1/4}$, this expression is bounded below by $(\lambda-\delta)^{\tau_1-2}\delta^{\tau_2-2}\,[\beta^2n^{3/2} - 2n]$. For fixed β and sufficiently large n, this is nonnegative, establishing the convexity of $\tilde{Q}$.

Now, to complete the proof, we essentially apply Jensen's inequality in the middle range. It is useful to consider two separate cases.

1. Suppose $\mathbf{Pr}[\text{small } \delta] = \mathbf{Pr}[\delta \leq p\lambda + n^{-1/4}] \geq 1/n$. Then restricting ourselves to this range, we have $\mathbf{E}[Q] \geq \frac{1}{n}\cdot 2^{-o(n)} = 2^{-o(n)}$.

2. On the other hand, suppose that the small values of δ have a probability at most $1/n$. Conditioning on the event that δ is in the middle or high range (i.e., $\delta \geq \lambda p + n^{-1/4}$), we have $\mathbf{E}[\delta \,|\, \text{middle or high range}] \leq \frac{\lambda(1-\lambda)}{1-\frac{1}{n}} = \lambda(1-\lambda) + o(1)$. Now, further conditioning on the middle range, the expectation can only go lower. That is,

   $$\mathbf{E}[\delta \mid \text{ middle range}] \leq \mathbf{E}[\delta \mid \text{ middle or high range}] \leq \lambda(1-\lambda) + o(1).$$

   Moreover, the probability of the middle range is at least $\lambda/2 - o(1) \geq \lambda/4$.

Therefore,

$$\mathbf{E}[Q] \geq (\lambda/4)\mathbf{E}[Q(\delta) \,|\, \text{middle range}] \geq (\lambda/4)\mathbf{E}[\tilde{Q}(\delta) \,|\, \text{middle range}].$$

Applying Jensen to the convex function $\tilde{Q}$, we get

$$\mathbf{E}[Q] \geq (\lambda/4)\tilde{Q}(\mathbf{E}[\delta \,|\, \text{middle range}]) \geq (\lambda/4)\tilde{Q}(\lambda(1-\lambda) + o(1)).$$

The final inequality uses the monotonicity of $\tilde{Q}$ and the fact that the conditional expectation of $\tilde{Q}$ is at most $\lambda(1-\lambda) + o(1)$. Finally, it remains to estimate $Q(\lambda(1-\lambda) + o(1))$. Plugging in $\lambda(1-\lambda)$ in place of δ, we see that

$$(1-p)\log(\lambda - \delta) + p\log\delta + h(p) - h(\lambda) = (1-p)\log(\lambda^2) + p\log(\lambda(1-\lambda)) + h(p) - h(\lambda),$$

which on rearranging equals $A := (1-p)\log\left(\frac{1-p}{\lambda}\right) + p\log\left(\frac{p}{1-\lambda}\right)$. Therefore, $\tilde{Q}(\lambda(1-\lambda) + o(1)) \geq 2^{-(A\beta + O_p(\beta^2))n}$.

Therefore, $\mathbf{E}[Q]$ is at least the minimum of the two estimates, which is $\geq 2^{-(A\beta + B\beta^2)n}$. ■

Claim 22. *Suppose* $A := (1-p)\log\left(\frac{1-p}{\lambda}\right) + p\log\left(\frac{p}{1-\lambda}\right)$. *Then,*

$$A - (1-2p)\frac{(h(\lambda) - h(p))}{\lambda - p} \leq \frac{5(\lambda - p)}{p}.$$

Proof: Applying the inequality $\ln z \leq z - 1$ with $z = \frac{1-p}{1-\lambda}$, we get

$$\log(\frac{1}{1-\lambda}) \quad \leq \quad \log(\frac{1}{1-p}) + (\lambda - p)\frac{\log e}{(1-p)} \leq \log(\frac{1}{1-p}) + 4(\lambda - p).$$

since $p < 1/2$ and $e < 4$. Plugging this in the definition of A, and also using $\lambda \geq p$, we get

$$A \leq (1-2p)\log(\frac{1-p}{p}) + 4p(\lambda - p) \leq (1-2p)h'(p) + 2(\lambda - p).$$

On the other hand, by the Lagrange Mean Value Theorem, there exists $\xi \in (p, \lambda)$ such that $(h(\lambda) - h(p)) = h'(\xi)(\lambda - p)$. Since h' is monotonically decreasing in $(0, 1/2)$, we have

$$\frac{(h(\lambda) - h(p))}{\lambda - p} = h'(\xi) \geq h'(\lambda).$$

Finally, we have

$$h'(\lambda) = h'(p) - \int_p^\lambda |h''(z)|\mathrm{d}z = h'(p) - \int_p^\lambda \frac{\log e}{z(1-z)}\mathrm{d}z \geq h'(p) - \frac{2(\lambda - p)}{p(1-p)} \geq h'(p) - \frac{4}{p}(\lambda - p),$$

again using $e < 4$.

Plugging in both these estimates, we get

$$\begin{aligned} A - (1-2p)\frac{(h(\lambda) - h(p))}{\lambda - p} &\leq (1-2p)h'(p) + 2(\lambda - p) - (1-2p)h'(p) + \frac{4(1-2p)}{p}(\lambda - p) \\ &\leq (2 + \frac{4(1-2p)}{p})(\lambda - p). \end{aligned}$$

Finally, using the obvious inequalities $2 < 1/p$ and $4(1-2p)/p < 4/p$, we get the result. ■

Claim 23. *Suppose* $A := (1-p)\log\left(\frac{1-p}{\lambda}\right) + p\log\left(\frac{p}{1-\lambda}\right)$ *and* $B = B(p) < \infty$. *Let* ϵ *be sufficiently small and* $\beta := (\lambda - p)/(1 - 2p + \epsilon)$. *Then*

$$A\beta + B\beta^2 \leq h(\lambda) - h(p) - A_0(\lambda - p)\epsilon + B_0(\lambda - p)^2$$

for some $A_0 > 0$ *and* $B_0 < \infty$ *depending on* p *(and independent of* ϵ *and* λ*).*

Proof: From Claim 22, we have

$$\begin{aligned} A\beta &\leq \left[\frac{1-2p}{\lambda - p}(h(\lambda) - h(p)) + \frac{5(\lambda - p)}{p}\right] \cdot \frac{\lambda - p}{1 - 2p + \epsilon} \\ &= \frac{1-2p}{1-2p+\epsilon}(h(\lambda) - h(p)) + \frac{5(\lambda - p)^2}{p(1-2p)} \end{aligned}$$

Assuming $\epsilon < 1 - 2p$, we can upper bound this by

$$\begin{aligned} A\beta &\leq \frac{1-2p-\epsilon/2}{1-2p}(h(\lambda) - h(p)) + \frac{5(\lambda - p)^2}{p(1-2p)} \\ &= h(\lambda) - h(p) - \frac{\epsilon}{2(1-2p)}(h(\lambda) - h(p)) + \frac{5(\lambda - p)^2}{p(1-2p)} \end{aligned}$$

Now, by the convexity of $h(\cdot)$, we have

$$\frac{h(\lambda) - h(p)}{\lambda - p} \geq \frac{h(1/2) - h(p)}{1/2 - p} = \frac{2(1 - h(p))}{1 - 2p}.$$

Therefore, we have

$$\begin{aligned} A\beta &\leq h(\lambda) - h(p) - \frac{\epsilon}{2(1-2p)}\frac{2(\lambda - p)(1 - h(p))}{1-2p} + \frac{5(\lambda - p)^2}{p(1-2p)} \\ &\leq h(\lambda) - h(p) - \epsilon(\lambda - p)\frac{1 - h(p)}{(1-2p)^2} + \frac{5(\lambda - p)^2}{p(1-2p)}. \end{aligned}$$

Also, $B\beta^2 \leq \frac{B(\lambda-p)^2}{(1-2p)^2}$. Therefore,

$$A\beta + B\beta^2 \leq h(\lambda) - h(p) - \epsilon(\lambda - p)\frac{1 - h(p)}{(1-2p)^2} + (\frac{5}{p(1-2p)} + \frac{B}{(1-2p)^2})(\lambda - p)^2.$$

Therefore the claim holds with $A_0 := \frac{1-h(p)}{(1-2p)^2}$ and $B_0 := \frac{5}{p(1-2p)} + \frac{B}{(1-2p)^2}$. ■

B Bounds for Random codes (Erasure case)

We recall the notation. Let $[q]_* := [q] \cup \{*\}$. For $a \in [q]_*^n$, define $\mathrm{supp}^*(a)$ to be the set of all indices i such that $a_i \neq *$. Let $\mathcal{E}_q(n, r)$ be the set of $a \in [q]_*^n$ such that $|\mathrm{supp}^*(a)| = n - r$. (We have $|\mathcal{E}_q(n, r)| = \binom{n}{r}q^{n-r}$.) Say that $a, b \in [q]_*^n$ agree with each other if $a_i = b_i$ for all $i \in \mathrm{supp}^*(a) \cap \mathrm{supp}^*(b)$. Finally, we will abbreviate $1 - p$ by α and $\mathcal{E}_q(n, pn)$ by $\mathcal{E}$.

B.1 Random General codes

Theorem 24 (Theorem 20 restated). *For every $0 < p < 1$ and $\gamma > 0$, with probability $1 - 2^{-\Omega_{p,\gamma}(n)}$, a random code of blocklength n and rate at least $1 - p - \gamma$ is not $(p, \frac{1-p}{2\gamma})$-erasure list decodable.*

Proof: Consider a random code of blocklength n and size 2^k, where $k = (\alpha - \gamma)n$, where $\alpha = 1 - p$. For a list of L messages $X = \{x_1, x_2, \ldots, x_L\} \subseteq \{0,1\}^k$, and $a \in \mathcal{E}$, define the indicator random variable $I(X, a)$ to be 1 iff $\mathbf{Enc}(x)$ agrees with a for all $x \in X$. Let $\mathcal{Z} := \sum_{X,a} I(X, a)$. It is clear that the code C is (p, L)-erasure list decodable if and only if $\mathcal{Z} = 0$.

For every X and a, we have $\mathbf{Pr}[I(X, a) = 1] = q^{-\alpha Ln}$, so that $\mathbf{E}[\mathcal{Z}] = q^{-\alpha Ln}\binom{q^k}{L}\binom{n}{np}q^{\alpha n} \geq L^{-L}q^{-(\alpha n - k)L}q^{\alpha n}\binom{n}{np}^2$ using standard approximations. Also,

$$\mathbf{Var}[\mathcal{Z}] \leq \sum_{X \cap Y \neq \emptyset} \sum_{a,b} \mathbf{Pr}[I(X, a) = 1 \text{ and } I(Y, b) = 1].$$

Fix an arbitrary pair (X, Y) with $|X \cap Y| = \ell > 0$. Further, let S, T denote the supports of a and b respectively. Now, suppose $I(X, a) = I(Y, b) = 1$. Then, for an arbitrary $z \in X \cap Y$, $\mathbf{Enc}(z)$ agrees with both a and b. Since $\mathbf{Enc}(z)$ is a string over $\{0, 1\}$ (not involving $*$), this implies that a, b must themselves agree with each other.

The event $I(X, a) = I(Y, b) = 1$ requires that the encodings of points in $X \backslash Y$ (resp., $Y \backslash X$) agree with a (resp. b), whereas for $z \in X \cap Y$, $\mathbf{Enc}(z)$ must agree with *both* a, b. Therefore, the probability of this event is at most

$$q^{-(|S||X\backslash Y| + |T||Y\backslash X|)}q^{-|S\cup T||X\cap Y|} = q^{-2\alpha(L-\ell)n}q^{-|S\cup T|\ell}$$

Summing over all pairs (a, b), and noting that the number of pairs (a, b) such that $\mathrm{supp}^*(a) = S$, $\mathrm{supp}^*(b) = T$, and $a|_{S\cap T} = b|_{S\cap T}$ is equal to $q^{|S\cup T|}$, we get

$$\begin{aligned}
\sum_{a,b} \mathbf{Pr}[I(X, a) = 1 \text{ and } I(Y, b) = 1] &= \sum_{S,T} q^{-2\alpha(L-\ell)n}q^{-|S\cup T|\ell}q^{|S\cup T|} \\
&\leq \binom{n}{\alpha n}^2 q^{-2\alpha(L-\ell)n}q^{-\alpha n(\ell-1)} \\
&= \binom{n}{pn}^2 q^{-\alpha(2L-\ell-1)n}
\end{aligned}$$

Finally, summing over X, Y pairs with $X \cap Y \neq \emptyset$, we get

$$\begin{aligned}
\mathbf{Var}[\mathcal{Z}] &\leq \sum_{\ell=1}^{L} L^{2L}q^{k(2L-\ell)}q^{-\alpha n(2L-\ell)}q^{\alpha n}\binom{n}{np}^2. \\
&\leq \sum_{\ell=1}^{L} L^{2L}(q^{-k}q^{\alpha n})^{\ell}q^{-\alpha n}\mathbf{E}[\mathcal{Z}]^2. \\
&\leq L^{2L+1}q^{(\gamma L-\alpha)n}\mathbf{E}[\mathcal{Z}]^2.
\end{aligned}$$

Therefore, for $L = \alpha/2\gamma$, we have $\mathbf{Var}[\mathcal{Z}] = q^{-\Omega_{p,\gamma}(n)}\mathbf{E}[\mathcal{Z}]^2$, and we are done. ■

B.2 Random Linear codes

Theorem 25 (Theorem 21 restated). *There exists a constant $c_q > 0$ such that for every $0 < p < 1$ and $\gamma > 0$, with probability $1 - q^{-\Omega_{p,\gamma}(n)}$, a random linear code of rate at least $1 - p - \gamma$ is not $(p, \frac{1}{2}q^{\frac{c_q p(1-p)}{2\gamma}})$-erasure list decodable with high probability.*

Proof: First of all, note that if we demonstrate a bad list containing L *linearly independent* points, then it automatically implies a general bad list of size q^{L-1}. This follows from the fact that if $c_1, c_2, \ldots, c_L$ agree with a, then any linear combination $\zeta_1 c_1 + \ldots + \zeta_L c_L$ also agrees with a, as long as $\zeta_1 + \zeta_2 + \ldots + \zeta_L = 1$. (Note that the number of such linear combinations is exactly q^{L-1}.)

Consider a random *linear* code C of dimension $k = (\alpha - \gamma)n$, where $\alpha = 1 - p$. For a *linearly independent* set of L messages $X \subseteq \{0,1\}^k$, and for every $a \in \mathcal{E}_q(n, pn)$, define $I(X, a)$ to be the indicator random variable for the event that $\mathbf{Enc}(x)$ agrees with a for all $x \in X$. Also, define Z to be $\sum_{X,a} I(X, a)$. For a fixed X and a, $\mathbf{E}[I(X, a)] = q^{-\alpha L n}$. Summing over the linearly independent L-tuples X, we get

$$\mathbf{E}[\mathcal{Z}] \geq \frac{1}{2} L^{-L} q^{kL} \cdot q^{-\alpha L n} \cdot q^{\alpha n} \binom{n}{np}$$

Define $\ell = \ell(X, Y) := \dim(\mathrm{span}(X) \cap \mathrm{span}(Y))$. For a pair X, Y of lists, say $X \sim Y$ if $\mathrm{span}(X)$ and $\mathrm{span}(Y)$ have a nontrivial intersection; that is, $\ell > 0$. If $X \not\sim Y$, then X and Y are linearly independent of each other. In turn, the random variables $I(X, a)$ and $I(Y, b)$ are also independent of each other. So, we get

$$\mathbf{Var}[\mathcal{Z}] \leq \sum_{X \sim Y} \sum_{a,b} \mathbf{Pr}[I(X, a) = 1 \text{ and } I(Y, b) = 1]$$

Fix a pair X, Y such that $\dim(\mathrm{span}\ X \cap \mathrm{span}\ Y) = \ell > 0$. As in Theorem 18, we define Z and $y_0 \in Y \backslash Z$ and write $y_0 = \sum_{u \in X \cup Z} \zeta(u) u$ for some scalars $\zeta(u)$. For any $a, b \in \mathcal{E}$, let $S = \mathrm{supp}^*(a)$ and $T = \mathrm{supp}^*(b)$. (Note that for general codes, for the event $I(X, a) = I(Y, b) = 1$ to occur, the strings a and b had to agree with each other on $S \cap T$; this is not so for linear codes.) For any $x \in X$, conditioned on the event $\mathbf{Enc}(x)|_S = a|_S$, the random variable $\mathbf{Enc}(x)|_{T \backslash S}$ is uniformly distributed over $\{0,1\}^{|T \backslash S|}$. Since $y_0 = \sum_{x \in X} \zeta(x)\, x + \sum_{z \in Z} \zeta(z)\, z$ with $\zeta(x) \neq 0$ for some $x \in X$, it follows that $\mathbf{Enc}(y_0)|_{T \backslash S}$ is also uniformly distributed over $\{0,1\}^{|T \backslash S|}$. Hence, conditioned on the event that $\mathbf{Enc}(x)$ agrees with a for all $x \in X$ and $\mathbf{Enc}(z)$ agrees with b for all $z \in Z$, the probability that $\mathbf{Enc}(y_0)$ agrees with b is at most $q^{-|T \backslash S|}$. Hence,

$$\begin{aligned} \sum_{a,b} \mathbf{Pr}[I(X, a) = I(Y, b) = 1] &\leq q^{-\alpha n(2L - \ell)} \cdot \sum_{S,T} q^{|S| + |T|} q^{-|T \backslash S|}, \\ &= q^{\alpha n(\ell - 2L)} q^{2\alpha n} \binom{n}{np}^2 \mathbf{E}_{S,T}\left[q^{-|T \backslash S|}\right] \end{aligned}$$

where the expectation is over $S, T \subseteq [n]$ of size $(1 - p)n$, chosen independently and uniformly randomly. By Lemma 26,

$$\sum_{a,b} \mathbf{Pr}[I(X, a) = I(Y, b) = 1] \leq q^{-\alpha n(2L - \ell)} q^{2\alpha n} \binom{n}{np}^2 q^{-c_q p(1-p)n}.$$

for some $c > 0$. The variance of $\mathcal{Z}$ can thus be bounded by

$$\begin{aligned}
\mathbf{Var}[\mathcal{Z}] &\leq \sum_{\ell=1}^{L} L^{2L} q^{k(2L-\ell)} q^{\alpha n(\ell-2L)} q^{2\alpha n} \binom{n}{np}^2 q^{-c_q p(1-p)n} \\
&\leq L^{2L} \sum_{\ell=1}^{L} q^{\gamma \ell n} q^{-c_q p(1-p)n} \left(q^{(k-\alpha n)L} q^{\alpha n} \binom{n}{np} \right)^2 \\
&\leq L^{4L+1} q^{\gamma L n} q^{-c_q p(1-p)n} (\mathbf{E}\mathcal{Z})^2
\end{aligned}$$

where the summand is maximized again for $\ell = L$. For $k = (\alpha - \gamma)n$, letting $L = (c_q/2) \cdot p(1-p)/\gamma$, we have $\mathbf{Var}\ \mathcal{Z} = q^{-\Omega(n)} (\mathbf{E}\mathcal{Z})^2$. We are thus done by an application of the second moment method. ■

Lemma 26. *There exists $c_q > 0$ (independent of n and p) such that if S, T are independent random subsets of $[n]$ of size $(1-p)n$, then*

$$\mathbf{E}_{S,T} \left[q^{-|T \backslash S|} \right] = O\left(q^{-c_q p(1-p)n} \right).$$

Proof: We prove this by thresholding on the value of $|T \backslash S|$. By symmetry, the quantity $\mathbf{E}_{S,T} \left[q^{-|T \backslash S|} \right]$ is the same as $\mathbf{E}_T \left[q^{-|T \backslash S|} \right]$ where S is fixed to be $\{1, 2, \ldots, (1-p)n\}$. In this case, the random variable $|T \cap S|$ has the hypergeometric distribution with mean $(1-p)^2 n$. We will first upper bound the probability of the event that $|T \backslash S| \leq \frac{1}{2} p(1-p)n$, which is equivalent to the tail event $|T \cap S| \geq \mathbf{E}[|T \cap S|] + \frac{1}{2} p(1-p)n$. By a standard Hoeffding bound for hypergeometric variables,

$$\begin{aligned}
\mathbf{Pr}\left[|T \backslash S| \leq \frac{1}{2} p(1-p)n \right] &= \mathbf{Pr}\left[|T \cap S| \geq (1-p)^2 n + \frac{1}{2} p \cdot (1-p)n \right] \\
&\leq \left(\left(\frac{1-p}{1-p/2} \right)^{1-p/2} \left(\frac{p}{p-p/2} \right)^{p-p/2} \right)^n \\
&= 2^{-\left((1-p/2) \log \frac{1-p/2}{1-p} - p/2 \right) n}
\end{aligned}$$

It can be checked that in the interval $[0, 1)$, the inequality

$$(1-p/2) \log \frac{1-p/2}{1-p} \geq 7p/10$$

holds, so that the tail probability is given by $\mathbf{Pr}\left[|T \backslash S| \leq \frac{1}{2} p(1-p)n \right] \leq 2^{-pn/5}$. Finally, the expectation is bounded as

$$\begin{aligned}
\mathbf{E}\left[2^{-|T \backslash S|} \right] &\leq \mathbf{Pr}\left[|T \backslash S| \leq \frac{1}{2} p(1-p)n \right] \cdot 1 + \mathbf{Pr}\left[|T \backslash S| \geq \frac{1}{2} p(1-p)n \right] \cdot q^{-\frac{1}{2} p(1-p)n} \\
&\leq 2^{-pn/5} + q^{-\frac{1}{2} p(1-p)n} \leq 2 \cdot q^{-c_q p(1-p)n}
\end{aligned}$$

for $c_q = 1/(5 \log q)$. ■

ECCC ISSN 1433-8092
http://eccc.hpi-web.de