Generalized List Decoding

Yihan Zhang
Department of Information Engineering, The Chinese University of Hong Kong
https://sites.google.com/view/yihan/
zy417@ie.cuhk.edu.hk

Amitalok J. Budkuley
Department of Electronics and Electrical Communication Engineering,
Indian Institute of Technology Kharagpur, India
http://www.facweb.iitkgp.ac.in/~amitalok
amitalok@ece.iitkgp.ac.in

Sidharth Jaggi
Department of Information Engineering, The Chinese University of Hong Kong
https://scholar.google.com.hk/citations?user=AX7276AAAAAJ&sortby=pubdate
jaggi@ie.cuhk.edu.hk

Abstract

This paper concerns itself with the question of list decoding for *general adversarial channels*, e.g., bit-flip (XOR) channels, erasure channels, AND (Z-) channels, OR (Σ-) channels, real adder channels, noisy typewriter channels, etc. We precisely *characterize* when exponential-sized (or positive *rate*) $(L-1)$-*list decodable* codes (where the *list size* L is a universal constant) exist for such channels. Our criterion essentially asserts that:

> For any given general adversarial channel, it is possible to construct positive rate $(L-1)$-list decodable codes *if and only if* the set of *completely positive tensors* of order-L with admissible marginals is not entirely contained in the order-L *confusability set* associated to the channel.

The sufficiency is shown via random code construction (combined with expurgation or time-sharing). The necessity is shown by

1. extracting approximately equicoupled subcodes (generalization of equidistant codes) from *any* sequence of "large" codes using hypergraph Ramsey's theorem, and
2. significantly extending the classic *Plotkin bound* in coding theory to list decoding for general channels using duality between the completely positive tensor cone and the *copositive* tensor cone.

In the proof, we also obtain a new fact regarding asymmetry of joint distributions, which may be of independent interest.

Other results include

1. List decoding capacity with asymptotically large L for general adversarial channels;
2. A *tight* list size bound for *most constant composition* codes (generalization of constant weight codes);
3. Rederivation and demystification of Blinovsky's [9] characterization of the list decoding *Plotkin points* (threshold at which large codes are impossible) for bit-flip channels;
4. Evaluation of general bounds ([43]) for *unique decoding* in the error correction code setting.

2012 ACM Subject Classification Mathematics of computing → Coding theory; Mathematics of computing → Information theory

Keywords and phrases Generalized Plotkin bound, general adversarial channels, equicoupled codes, random coding, completely positive tensors, copositive tensors, hypergraph Ramsey theory

Digital Object Identifier 10.4230/LIPIcs.ITCS.2020.51

Related Version An extended version of the paper will be updated at https://arxiv.org/abs/1909.04264.

Acknowledgements We thank Andrej Bogdanov who provided an elegant reduction from general L to $L = 2$ for the proof of the asymmetric case of the converse (Lemma 68) and rederived Blinovsky's [9] characterization of the Plotkin point P_{L-1} for $(p, L-1)$-list decoding via a conceptually cleaner proof (Sec. 16), despite that he generously declined to co-author this paper. We also thank him for inspiring discussions in the early stage and helpful comments near the end of this work.
Part of this work was done while YZ was visiting the Simons Institute for the Theory of Computing for the Summer Cluster: Error-Correcting Codes and High-Dimensional Expansion, and AJB was at the Department of Information Engineering, the Chinese University of Hong Kong. This work was partially supported by GRF grants 14301519 and 14313116.

1 Warmup

In favour of motivating general problems, introducing general notions and stating our general theorems, we first go through concrete numerical examples that are special cases of our results.

Suppose Alice can transmit a length-n bit string (*codeword*) to Bob and an adversary James can flip np $(0 \leqslant p \leqslant 1)$ of these bits. Consider first the classic coding theory question.

1. **Error correction.** For what values of p, can one construct a *code* (collection of codewords) of *positive rate* (i.e., codebook size at least 2^{Rn} for some constant $1 \geqslant R > 0$) such that Bob can uniquely decode? The classic Plotkin bound [33] tells us that this is impossible for $p > 1/4$,[1] and the classic Gilbert–Varshamov (GV) bound [22, 42] tells us that this is possible for $p < 1/4$.
2. **List decoding.** For what values of p, can one construct a code of positive rate such that it is *3-list decodable* (i.e., regardless of which np bits James flips, Bob can always decode the received word to a *list* of at most 3 codewords, one of which is the codeword transmitted by Alice)?[2] Due to work by Blinovsky, it is known that this is possible if and only if $p \leqslant 5/16$.[3]

In this work, not only are we are able to rederive all the above thresholds, but are also able to derive the corresponding thresholds for a vast variety of *general adversarial channels* such as bit-flip channels, erasure channels, AND (Z-) channels, OR (Σ-) channels, adder channels, noisy typewriter channels, etc..

In this section, let us revisit the answers to questions 1 and 2 in the technical language we develop in this paper.

1. **Error correction.** Consider any pair of codewords $\underline{x}_1, \underline{x}_2$ that are resilient to np bit-flips. They must therefore be at a Hamming distance larger than $2np$. Said differently, the *joint type* (i.e., the 2×2 matrix whose (x_1, x_2)-th, $x_1, x_2 \in \{0, 1\}$, entry is the fraction of locations i of $(\underline{x}_1, \underline{x}_2)$ such that $\underline{x}_1(i) = x_1$ and $\underline{x}_2(i) = x_2$) $\tau_{\underline{x}_1, \underline{x}_2} = \begin{bmatrix} t(0,0) & t(0,1) \\ t(1,0) & t(1,1) \end{bmatrix}$ of these two codewords must satisfy the condition that

   C1 $t(0,1) + t(1,0) \geqslant 2p$.

[1] Actually for $p = 1/4$ this is still impossible.
[2] Note that a 1-list decodable code is exactly a uniquely decodable code (or more commonly called an *error correction code*).
[3] In fact Blinovsky identified the threshold p up to which positive rate $(p, L-1)$-list decodable codes exist for *any* integer $L \geqslant 2$. This, in particular, recovers the Plotkin bound.

a. In [9, 34, 2][4] and [43], it was shown that: if a code $\mathcal{C}$ of size 2^{Rn} exists, then there must exist a *positive rate* subcode $\mathcal{C}' \subset \mathcal{C}$ such that for *every* pair of codewords $\underline{x}_1, \underline{x}_2$ in $\mathcal{C}'$, their joint type is approximately the same (as, say, $P_{\mathbf{x}_1,\mathbf{x}_2}$).

b. In [43], it was shown that: it is possible to construct positive rate codes with joint types (close to) $P_{\mathbf{x}_1,\mathbf{x}_2}$ if and only if $P_{\mathbf{x}_1,\mathbf{x}_2}$ is a *completely positive (*CP*)* distribution, i.e., $P_{\mathbf{x}_1,\mathbf{x}_2}$ can be written as a convex combination of products of independent and identical distributions,

$$P_{\mathbf{x}_1,\mathbf{x}_2} = \sum_{i=1}^{k} \lambda_i P_{\mathbf{x}_i} P_{\mathbf{x}_i}^{\top},$$

for some positive integer k, convex combination coefficients $\{\lambda_i\}_{1\leqslant i\leqslant k}$ and probability vectors $\{P_{\mathbf{x}_i}\}_{1\leqslant i\leqslant k}$. For example,

$$\lambda \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix} + (1-\lambda) \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix} \tag{1}$$

is CP for $\lambda \in [0,1]$ since it can be written as $\frac{\lambda}{2}\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}1 & 0\end{bmatrix} + \frac{\lambda}{2}\begin{bmatrix}0\\1\end{bmatrix}\begin{bmatrix}0 & 1\end{bmatrix} + (1-\lambda)\begin{bmatrix}1/2\\1/2\end{bmatrix}\begin{bmatrix}1/2 & 1/2\end{bmatrix}$. One can check that for $\lambda < 0$, matrix (1) is not CP. For condition **C1** to be satisfied by some CP distribution, it must be the case that $2p \leqslant 2\cdot(1-\lambda)\cdot(1/4)$ for some $\lambda \in [0,1]$. This is impossible if $p > 1/4$. As a consequence, the classic Plotkin bound is recovered in this convex geometry language, since the non-CP matrices of the form (1) with *negative* λ correspond to codes with minimum pairwise fractional distance $\frac{1+|\lambda|}{2}$ (hence correspond to $p = \frac{1+|\lambda|}{4} > 1/4$), which, by the Plotkin bound, cannot have positive rate.

2. **List decoding.** Now let us move to the list decoding question in hands. For a code to be 3-list decodable, it must be the case that for any quadruple $\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_4$, there is no $\underline{y}$ such that the Hamming distance from $\underline{x}_i$ to $\underline{y}$ is at most np for *every* $i \in \{1,2,3,4\}$. In this case, the appropriate object is therefore a $2\times2\times2\times2$ tensor (or a joint distribution of $(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4)$) $P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4}$ such that

C2 any of its *extension* $P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4,\mathbf{y}}$ (i.e., a coupling of $(\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4)$ and $\mathbf{y}$, or a $2\times2\times2\times2\times2$ tensor such that $P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4} = P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4,\mathbf{y}=0} + P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4,\mathbf{y}=1}$) satisfies the condition that $P_{\mathbf{x}_i,\mathbf{y}}(0,1) + P_{\mathbf{x}_i,\mathbf{y}}(1,0) > p$ for at least one $i \in \{1,2,3,4\}$.

a. Again, by [9, 34, 2] and our work, we can restrict our attention to codes in which every 4-tuple of codewords has joint type close to some $P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4}$, since we can find such a subcode which is sufficiently large in *any* positive rate code.

b. Generalizing [43], we show that codes with order-4 joint types (close to) $P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4}$ exist if and only if $P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4}$ is a *completely positive tensor* of order-4, i.e., $P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4}$ can be written as a convex combination of products of independent and identical distributions,

$$P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{x}_4} = \sum_{i=1}^{k} \lambda_i P_{\mathbf{x}_i}^{\otimes 4}.$$

[4] Their and our work showed that it is also possible to find a positive rate subcode such that every L-tuple of codewords has joint type close to some $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$. This, as we shall see momentarily, is useful for list decoding.

One can check that distributions of the form

$$\lambda \operatorname{diag}(1/2) + (1-\lambda)\begin{bmatrix}1/2\\1/2\end{bmatrix}^{\otimes 4} = \frac{\lambda}{2}\begin{bmatrix}1\\0\end{bmatrix}^{\otimes 4} + \frac{\lambda}{2}\begin{bmatrix}0\\1\end{bmatrix}^{\otimes 4} + (1-\lambda)\begin{bmatrix}1/2\\1/2\end{bmatrix}^{\otimes 4}$$

is CP if and only if $\lambda \in [0,1]$. On the other hand, for condition **C2** to be satisfied by some tensor like that, it turns out, as shown by Blinovsky [9] and us, that p has to be no larger than 5/16.

Of course, bit-flips are just one of the simplest models of corruption that may occur in real-world communication/storage systems. Perhaps, under certain circumstances, in the system of interest, we are allowed to transmit length-n codewords taking values from $\{0,1,2,3,4,5\}$, but each legitimate codeword $\underline{x}$ has to satisfy the following constraints inherently associated to this communication system

$$\begin{cases} \tau_{\underline{x}}(1) + 3\tau_{\underline{x}}(3) \leqslant 1.2 \\ \tau_{\underline{x}}(2) - \tau_{\underline{x}}(3) \geqslant 0.05 \\ \tau_{\underline{x}}(0) - \tau_{\underline{x}}(4) - 0.2\tau_{\underline{x}}(5) \leqslant 0.7 \end{cases},$$

where $\tau_{\underline{x}}(x)$ denotes the fraction of $x \in \{0,1,\cdots,5\}$ in $\underline{x} \in \{0,1,\cdots,5\}^n$. An adversary is allowed to change symbols in the transmitted codeword only from small values to large values; the cost he pays by changing every i to j $(0 \leqslant i < j \leqslant 5)$ is $j-i$ dollars, and he has a budget of $2.3n$ dollars in total. Among others, one of the fundamental questions we are able to answer in this paper is the following: is it possible for us to design exponentially large codes such that no matter which codeword is transmitted and no matter how an adversary corrupts it via a legitimate action, the decoder is always able to output a list of at most (say) 10 codewords which contains the correct one?

The answer to the above question is affirmative and can be stated in a similar manner: it is possible if and only if there is a CP tensor of order 11 and dimension 6 which does not lie inside the *confusability set* determined by the channel. In particular, the confusability set is the set of joint distributions which fail to meet the conditions similar to **C1** or **C2** that are determined by the channel.

Our results tell us that if one only aims to search for exponentially large $(L-1)$-list decodable codes (instead of optimizing its size) for a given general adversarial channel, then it is *sufficient* (and obviously necessary) to restrict our attention to codes that are *chunk-wise random-like*. Such codes correspond to some CP distribution $\sum_{i=1}^{k}\lambda_i P_{\mathbf{x}_i}^{\otimes L}$. If a random code of positive rate, where the $\lambda_i n$ $(1 \leqslant i \leqslant k)$ components in the i-th chunk of each codeword are sampled from distribution $P_{\mathbf{x}_i}$, does not "work" with high probability (w.h.p.), then we can never find positive rate codes of any other form that "work" for the underlying channel.

By setting the *list size* $L-1=1$, results in [43] are recovered by our work.

2 Introduction

While the main contribution of this work is to strictly generalize notions that have been primarily studied for "Hamming metric" channels, before we precisely define general channels, let us reprise what is known for Hamming metric channels in this section.

2.1 Error correction and the Plotkin bound

The theory of error correction codes is about protecting data from errors. In classic coding theory, a code, say $\mathcal{C}$, is just a collection of binary *codewords* (which are usually just binary length-n sequences, where n is called the *blocklength*). The most well-studied error model is *bit-flip*. When a certain codeword is transmitted, an adversary can arbitrarily flip at most np $(0 \leqslant p \leqslant 1)$ bits. It is easy to see that two codewords are not *confusable* if and only if their Hamming distance (number of locations where they differ, denoted $d_{\mathrm{H}}(\cdot,\cdot)$) is at least $2np+1$. Let

$$d_{\min}(\mathcal{C}) := \min_{\substack{\underline{x},\underline{x}'\in\mathcal{C}\\ \underline{x}\neq\underline{x}'}} d_{\mathrm{H}}\left(\underline{x},\underline{x}'\right)$$

denote the minimum pairwise distance of codewords in $\mathcal{C}$. The goal is to *pack* as many codewords as possible in the Hamming space $\mathbb{F}_2^n$ while ensuring that the minimum distance is at least $2np+1$. By a simple volume argument (Gilbert–Varshamov (GV) bound [22, 42]), it is known that *exponentially many* such vectors can be packed when $p < 1/4$. The fundamental quantity that coding theorists are seeking when faced with any communication model is the largest *achievable* rate, i.e., *capacity*. The *rate* $R(\mathcal{C})$ of a code $\mathcal{C}$ is its normalized cardinality, i.e., $R(\mathcal{C}) := \frac{\log|\mathcal{C}|}{n}$. The capacity C measures, asymptotically as the blocklength grows, the largest fraction of bits (out of n) that can be reliably transmitted despite np adversarial bit-flips. C is formally defined as[5]

$$C := \limsup_{n\to\infty} \max_{\mathcal{C}\subset\mathbb{F}_2^n:\, d_{\min}(\mathcal{C})>2np} R(\mathcal{C}).$$

For the aforementioned bit-flip model, as said, the problem of finding the capacity can be also cast as determining the *sphere packing density*. This problem is notoriously difficult and is still open to date. However, we do know that $p = 1/4$ is the threshold below which exponential-sized packing exists (as suggested by the GV bound) and above which it is impossible. The latter fact is the famous Plotkin bound. More formally,

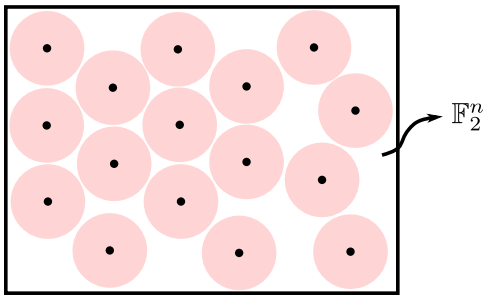
▶ **Theorem 1** (Plotkin bound [33]). *If $p = 1/4 + \epsilon$, $\epsilon > 0$, then any code $\mathcal{C}$ of distance larger than $2np$ has cardinality at most $1 + \frac{1}{4\epsilon}$ (and hence has zero rate).*

We will call the value of p at which the capacity hits zero, the *Plotkin point*. Note that the Plotkin bound actually tells us that above the Plotkin point, not only does every code/packing have a size $2^{o(n)}$ (and hence, rate zero), but also that its size should be at most a *constant* (independent of the blocklength n). Coupled with the achievability result given by the GV bound, the phase transition threshold for exponential-sized packing is thereby identified precisely.

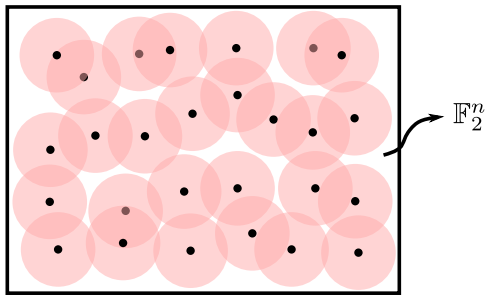
2.2 List decoding and the list decoding Plotkin bound

We now introduce another important notion: *list decoding*. List decodability still requires codewords to be separated out, but in a more relaxed sense; only a few codewords (instead of exactly one under unique decoding discussed earlier) can be captured by a ball of certain radius, no matter where it is located.

[5] It turns out that allowing vanishing probability of decoding error does not change the problem.



(a) An $(L-1)$-packing for $L=2$, i.e., disjoint packing.



(b) An $(L-1)$-packing for $L=3$, i.e., packing with multiplicity 2.

Figure 1 Packing (uniquely decodable codes) vs. multiple packing (list decodable codes). The geometry depicted in the above figures may be misleading compared with the truth in binary Hamming space.

▶ **Definition 2** (List decodability [21, 46]). *A code $\mathcal{C}$ is $(p, L-1)$-list decodable (or $(p, <L)$-list decodable) if for all $\underline{y} \in \mathbb{F}_2^n$, $\left|\mathcal{C} \cap \mathcal{B}_H\left(\underline{y}, np\right)\right| < L$, where $\mathcal{B}_H(\underline{y}, np)$ denotes a Hamming ball centered at $\underline{y}$ of radius np.*

Of course we want the *list size* L to be as small as possible. In particular, the problem is trivial when $L = |\mathcal{C}|$. (The decoder ignores the received word and outputs the whole code.) When $L = 2$, the problem precisely becomes packing. As the admissible L grows, the problem is expected to become easier.

Introduced by Elias [21], list decoding is an important and well-studied subject in coding theory. It is a natural mathematical question to pose towards understanding high-dimensional geometry in discrete spaces. It also serves as a primitive that is useful within and beyond the scope of coding theory. For instance, in many communication problems (e.g., [1, 13]), one proof technique is to let the decoder list decode to a short list (usually $\text{poly}(n)$-sized suffices) of candidate messages, then use other information to disambiguate the list and get a unique message. List decoding also finds applications in complexity theory, cryptography, etc. [25]. For instance, it is used for amplifying hardness and constructing extractors, pseudorandom generators and other pseudorandom objects [20]. The idea of relaxing the problem by asking the solver to just output a list (ideally as small as possible) of solutions that is guaranteed to contain the correct one, instead of insisting on a unique answer, is also adopted in many other fields in computer science [19, 35, 28]. In the context of high-dimensional geometry over finite fields, list decoding is equivalent to multiple packing, just like error correction codes are equivalent to sphere packing. Multiple packing is a natural generalization of the famous sphere packing problem in which, instead of insisting on disjoint alignment, overlap with bounded multiplicity is allowed.

▶ **Definition 3** (Multiple packing). *A subset $\mathcal{C} \subset \mathbb{F}_2^n$ is a $(p, L-1)$-multiple packing if when we put Hamming balls of radii np around each vector in $\mathcal{C}$, no point in the space simultaneously lies in the intersection of at least L balls.*

See Fig. 1 for examples of packing and multiple packing in the Hamming space.

Surprisingly, list decoding capacity is known if we allow L to be asymptotically large. In some sense, list decoding makes us information-theoretic since in many (though not all) cases the list decoding capacity coincides with the capacity of the corresponding Shannon channels in which the noise is random with the same "power" (e.g., in the bit-flip/bit-erasure case, each component of the random noise is independently and identically distributed (i.i.d.) according to a Bernoulli distribution with mean p).

▶ **Theorem 4** (List decoding capacity (folklore) [47]). *Given any $\delta > 0$, there exists an infinite sequence of $(p, \mathcal{O}(1/\delta))$-list decodable codes $\{\mathcal{C}_n\}_n$, each of rate $1 - H(p) - \delta$. Indeed, for any sufficiently large n, a random code (each codeword sampled uniformly at random from $\mathbb{F}_2^n$) of rate $1 - H(p) - \delta$ is $(p, \mathcal{O}(1/\delta))$-list decodable w.h.p..*

On the other hand, any infinite sequence of codes of rate $1 - H(p) + \delta$ is $\left(p, 2^{\Omega(n\delta)}\right)$-list decodable.

*We call $1 - H(p)$ the p-*list decoding capacity *(without specifying a specific L). In particular, the Plotkin point for p-list decoding when L is sufficiently large is $1/2$.*

Though the fundamental limit for the relaxed problem for large constant L is essentially understood, $(p, L-1)$-list decodability for small L (e.g., absolute constant, say $3, 8, 100$, etc.; or sublinear in $1/\delta$, say $(1/\delta)^{1/2}$, $(1/\delta)^{1/3}\log(1/\delta)$, $\log\log(1/\delta)$) is far from being understood. Indeed, it is believed (at least for absolute constant L) to be equivalently hard as the sphere packing problem. Formally, the question of understanding the role of L can be cast as follows. Note first that when $L = 2$, the (unknown) capacity lies somewhere between the Gilbert–Varshamov bound and the Linear Programming bound ([18, 30, 44, 31, 32]). When $L = \mathcal{O}(1/\delta)$, the list decoding capacity $1 - H(p)$ is much larger than the unique decoding capacity. As we increase L, the $(p, L-1)$-list decoding capacity should be gradually "lifted" and the corresponding Plotkin point (the value of p where the capacity is zero) should somehow move rightwards from $1/4$ to $1/2$. The principal goal is to completely understand the dynamics of this evolution.

▶ Remark 5. In this paper, we explicitly distinguish the list decoding capacity for large L and for small L. When we say that L is asymptotically large, we refer to $L = \Omega(1/\delta)$ which suffices to approach the p-list decoding capacity within gap δ. When we say that L is small without further specification, we refer to absolute constant L. For large L, the p-list decoding capacity, denoted by C (recall that we do not explicitly specify L for this regime, see Theorem 4), is fully characterized; however, the $(p, L-1)$-list decoding capacity for small L, denoted by C_{L-1}, is widely open.

Again, for any absolute constant L, the $(p, L-1)$-list decoding capacity is poorly understood. We only have non-matching lower and upper bounds. To our knowledge, the current best bounds are due to Blinovsky from the 80s [9, 10, 11], except for sporadic values of L in some regimes of p. Specifically, for $L = 3$, Ashikhmin–Barg–Litsyn [3] can uniformly improve Blinovsky's upper bound for all values of p. For *even* L's that are at least 4, Polyanskiy [34] can partially beat Blinovsky's upper bounds in the low rate regime.

Though how C_{L-1} approaches C as L increases is not exactly known, Blinovsky's bounds *do* resolve the dynamics of the Plotkin point evolution! Let P_{L-1} denote the Plotkin point for $(p, L-1)$-list decoding. Let $L = 2k$ or $2k+1$ ($k \geqslant 1$). Then Blinovsky's results imply that P_{L-1} is precisely given by the following formula

$$P_{L-1} = \sum_{i=1}^{k} \frac{\binom{2(i-1)}{i-1}}{i} 2^{-2i}.$$

Later, Alon–Bukh–Polyanskiy [2] recovered this result with a simpler-looking formula

$$P_{L-1} = \frac{1}{2} - 2^{-2k-1}\binom{2k}{k},$$

For instance, $P_1 = P_2 = 1/4$, $P_3 = P_4 = 5/16$, etc. As can be noted, the Plotkin point moves *periodically*! The fact that the above two formulas always evaluate to the same value is implicit in [2] and is formally established in Appendix D.

3 Our contributions

Our motivation comes from a well-known connection between list decodability and reliability of communication over adversarial channels. A binary code is $(p, L-1)$-list decodable if and only if it has zero error when used over the following *adversarial bit-flip channel* (Fig. 2).

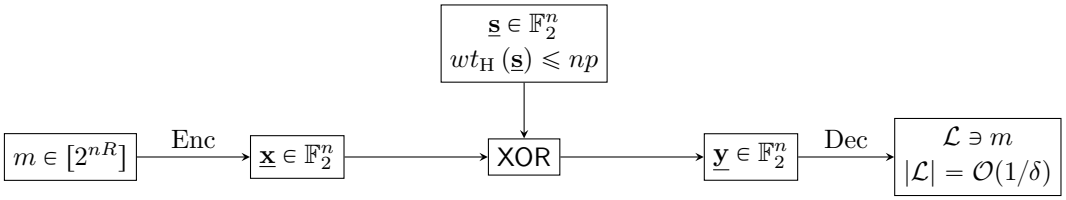


Figure 2 Adversarial bit-flip channels.

The above system depicts a one-way point-to-point communication scenario in which the encoder (Alice) randomly picks a message m from 2^{nR} of them and encodes it into an n-bit string. The adversary (James) stares at this entire codeword and maliciously flips at most np bits of it. Then, the decoder (Bob) receives the corrupted word and is required to output a short list of messages which is guaranteed to contain m with probability 1.

In the above model, the adversary is power constrained in the sense that he only has a budget of np bit-flips. But the encoder is not constrained – she can encode the message into any vector in $\mathbb{F}_2^n$. In some scenarios, codewords are also weight constrained. It makes sense to pose the same question (understanding the list decoding capacity) for input constrained channels. Indeed, this question was also studied in the literature [26].

Motivated by this connection, we significantly generalize the bit-flip model and define list decodability for *general adversarial channels*. We consider a large family of channels in which the encoder is allowed to encode the message into a length-n sequence $\underline{x}$ over *any* alphabet $\mathcal{X}$ of constant size, the adversary is allowed to design an adversarial noise pattern $\underline{s}$ over *any* alphabet $\mathcal{S}$ and the channel can be any *deterministic component-wise* function taking as input a pair of strings from $\mathcal{X}^n \times \mathcal{S}^n$, outputting a sequence $\underline{y}$ over *any* alphabet $\mathcal{Y}$ of the same length. The system designer can incorporate a large family of constraints on $\underline{x}$ and $\underline{s}$ in terms of their *types* (i.e., empirical distributions). The above family of adversarial channels we consider includes but is not limited to

1. The standard adversarial bit-flip channels and adversarial erasure channels;
2. Z-channels in which the adversary can only flip 1 to 0 but not the other way around;
3. Adder channels in which the output is the sum of inputs over the reals rather than modulo the input alphabet size;
4. Channels equipped with Lee distance instead of the Hamming metric.

Indeed, our framework covers most well-studied error models and more that potentially have not been studied in the literature.

However, since we require the channel transition function to act on each component of the inputs independently, a well-studied family of channels is excluded: the *adversarial deletion channels* (cf. [12]). In this model, the adversary can *delete* at most np entries of the transmitted codeword and the decoder receives a vector of smaller length (but at least $(1-p)n$) without knowing the original locations of the symbols he got[6]. Determining the

[6] We want to emphasize the difference between deletions and erasures. When symbols in the codeword are deleted, the rest of the symbols are concatenated and the receiver has no idea which symbols were

Plotkin point for this channel is a long standing open problem. It is known [12] that for binary channels the Plotkin point lies between $\sqrt{2}-1 \approx 0.414$ and 0.5; for q-ary channels, it lies between $1-\frac{2}{q+\sqrt{q}}$ and $1-\frac{1}{q}$. The capacity of this channel is even less understood.

For technical simplicity, we also assume that the channel transition function is *deterministic*, i.e., the output symbol y is a deterministic function of the codeword symbol x and the error symbol s.[7]

We can assume, without any loss of generality, that none of the encoder, decoder and adversary has private randomness to randomize their strategy. This is because there are reductions showing that, given stochastic encoder/decoder, we can construct a deterministic coding scheme with essentially the same rate. Similarly, given a stochastic adversarial error function, we can turn it into a deterministic one which is equivalently malicious in terms of rate. Therefore, for the encoder, it suffices to only consider deterministic codes where each message is mapped to a unique codeword with probability 1. For the adversary, we can assume the error pattern is a deterministic function of the transmitted codeword. Nevertheless, note that the error function does *not* have to be component-wise independent. The i-th component $\underline{s}(i)$ of the noise pattern $\underline{s}$ can depend on *every* entry of $\underline{x}$, not only on the corresponding $\underline{x}(i)$. Moreover, the decoder's decision on the estimated message given the received word can also be assumed to be deterministic. That is, we can require the decoder to output the correct message with *zero* error probability. Hence, the problem is purely combinatorial and all desirable events should happen with probability one.

In this work, we precisely *characterize* the Plotkin point for list decoding over any channel from the above family of general adversarial channels. That is, we essentially provide a criterion (sufficient and necessary condition) for the existence of positive rate $(L-1)$-list decodable codes for such channels.

In the context of high-dimensional geometry over finite fields, the result can be also cast as pinning down the location of the phase transition threshold for the optimal density of $(L-1)$-multiple packing using *general shapes* (not necessarily Hamming balls) corresponding to the defining constraints on codewords and errors of the channel. Above the threshold, exponential-sized multiple packing exists while below that, it is impossible to have such exponential-sized multiple packings.

This criterion can be summarized in one sentence:

> exponential-sized $(L-1)$-list decodable codes for general adversarial channels (or $(L-1)$-multiple packings using general shapes) essentially exist if and only if the *completely positive tensor cone* of order-L is not entirely contained in the *confusability set* of the channel for $(L-1)$-list decoding.

Jargon in the above informal statement will become understandable once we formalize the problem setup and present rigorous claims. The proof consists of the sufficiency part and the necessity part. At a very high level, the sufficiency part follows from a random coding argument and its generalization inspired by the time-sharing argument frequently used in Network Information Theory. The necessity part builds upon and significantly generalizes the classic Plotkin bound, which goes by first extracting an *equicoupled* subcode using Ramsey theory and then applying a generalized *double counting* trick.

---

deleted. However, when symbols are erased, they are replaced by erasure symbols erasure at the same locations and the receiver seeing them knows exactly which symbols were erased. Hence the erasure case is much simpler than the deletion case.

[7] The general case in which the channel law is given by a conditional distribution $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ (with not necessarily only singleton atoms) is more technical and is left as one of our future directions.

Our other results include the following:

1. For any given general adversarial channel, we pin down the list decoding capacity for asymptotically large L. This generalizes the classic list decoding capacity in the bit-flip case. The lower bound is achieved by a purely random code. The upper bound follows from a volume packing argument.
2. For any given general adversarial channel, we determine the *exact* order (in terms of δ) of the list sizes of a large fraction (exponentially close to one) of constant composition codes (in which all codewords have the same type) achieving the list decoding capacity within gap δ. It turns out that if we pick a constant composition code from the set of all such codes uniformly at random, with high probability, it is exactly $\Theta(1/\delta)$-list decodable.
3. For any given general adversarial channel and any $L \geqslant 2$, we give a lower bound on the $(L-1)$-list decoding capacity. It coincides with the generalized Gilbert–Varshamov bound obtained by [43] when $L-1$ is equal to 1. Our bound follows from a random code construction assisted by expurgation, generalizing a classic construction for $(p, L-1)$-list decoding in the bit-flip case [24]. Note that this construction differs from [43]'s construction for unique decoding using greedy packing.
4. In the special case where $L = 2$, i.e., the unique decoding setting and under the bit-flip model, we evaluate the Gilbert–Varshamov-type bound and an achievable rate expression of cloud codes (codes constructed from CP distributions) obtained by [43]. In particular, we show that the Gilbert–Varshamov-type bound for general adversarial channels matches the classic GV bound in coding theory. We also provide an explicit convex program for evaluating achievable rates of cloud codes.
5. By evaluating our general criterion under the bit-flip model, we *numerically* recover Blinovsky's [9] characterization of the Plotkin points for $(p, L-1)$-list decoding. This boils down to checking the feasibility of an explicit linear program with structured coefficient matrix. Though the LP has size exponential in L, its feasibility can be checked in constant time since our results are tailored for constant L independent of the blocklength n (which needs to approach infinity for many of our results to hold).
6. By utilizing facts discovered in this paper, we *rigorously* recover Blinovsky's [9] characterization of the Plotkin points for $(p, L-1)$-list decoding. Our proof avoids the complicated calculations Blinovsky did and demystifies the formula by Blinovsky[8]. In particular, our lower bound on the Plotkin point explains why, in the low rate regime, *average-radius*[9] list decoding is equivalent to the classic notion of list decoding. We believe that this fact was first observed and rigorously justified by Blinovsky. It was later rediscovered many times and became one of the basic starting points of many papers, especially those regarding list decoding random q-ary linear codes. Our upper bound relates the Plotkin point P_{L-1} to the expected translation distance of a one-dimensional unbiased random walk after L steps. In summary, using connections between codes and random variables, we are able to re-interpret the results by Blinvosky [9] and Alon–Bukh–Polyanskiy [2] within the framework we established by providing a more intuitive formula which matches known results.

[8] In fact, he provided upper and lower bounds on the $(p, L-1)$-list decoding capacity which happen to vanish at the same value of p.

[9] $(p, L-1)$-average-radius list decodability requires that the *average* distance (instead of maximum distance required by the classic notion of $(p, L-1)$-list decodability) from any L-tuple of codewords to their centroid is larger than np. Average-radius list decodability is a more stringent requirement since it implies the classic list-decodability. However, it is easier to analyze since the problem is *linearized* from infinity norm to one norm. Indeed it plays a useful role in a long line of work towards understanding the list decodability of random linear codes [26, 45, 36, 37, 38].

4 Overview of techniques

Our paper is highly correlated to a sister paper [43] which a subset of the authors are involved in. That paper provides generalized Plotkin bound for *unique* decoding over general adversarial channels. The authors showed that exponential-sized *uniquely* decodable codes or hard packings exist if and only if the set of completely positive *matrices* is not entirely contained in the *confusability set* associated to the given channel. This answers the question we posed in the beginning of the paper for the $L = 2$ case. We generalize their results to *any universal constant* L. Almost all results in [43] can be recovered by setting $L = 2$ in our paper.

We give an overview of the techniques used in this paper and highlight the similarities and differences between [43][10] and our work.

1. The general adversarial channel models that both papers are concerned with belong to a larger family of channels known as *Arbitrarily Varying Channels (AVC)* in Information Theory community; these were first studied by Blackwell et al. [8] (see [29] for a detailed survey). We want to emphasize that the bulk of the literature on AVCs deals with *oblivious* adversary channels in which the adversary has to pick his malicious noise pattern *before* the codeword is chosen from the codebook (and hence, *oblivious* of the transmitted codeword) by the encoder. This makes the problem significantly easier and the capacity of such channels is precisely known (cf. [17]). The channels that [43] and we are considering are such that the adversary gets to design the error pattern with the complete knowledge of the transmitted codeword; these are called *omniscient* adversaries in [43]. This problem is much more difficult and the capacity is, again, widely open even for simple models such as the bit-flip channels. Indeed, the subclass of AVCs that [43] and we defined is motivated by the bit-flip channels and its various variants, e.g., q-ary channels, weight constrained channels, asymmetric channels, etc..
2. The connection between codes and random variables/distributions is classic in Theoretical Computer Science. The idea of realizing binary error correction codes using $\{-1, 1\}$-valued random variables or functions supported on the Boolean hypercube $\{-1, 1\}^n$ is spread out in the literature explicitly or in disguise. Such a trick allows one to borrow tools from other fields of Theoretical Computer Science, e.g., the theory of expander graphs, randomness extractors, small-bias distributions, discrete Fourier analysis, etc., (cf. [40, 5, 41, 7]) to understand, construct and analyze codes.
3. With respect to (w.r.t.) codes for general adversarial channels, the specific idea of collecting admissible types of good codes and studying the set of corresponding distributions was used in [43]. In particular, they defined similar notions of self-couplings and confusability sets which are submanifolds of *matrices* corresponding to joint distributions. Such objects only take care of *pairwise* interaction of codewords, which is insufficient for understanding list decoding. We generalize their notions to *tensors* which capture the (empirical) joint distributions of *lists* of codewords. While some properties in [43] continue to hold when objects in the matrix case are extended to their tensor versions, others fail to hold, as we will see in the rest of the paper. We also encounter issues which do not arise in the unique decoding setting. As is well-known, tensors are much more delicate [27] to handle compared to matrices.

[10] Though the work by Wang–Budkuley–Bogdanov–Jaggi [43] has been accepted to ISIT 2019, the conference version is limited to 5 pages and contains essentially no proof. At the time this paper is written, we do not have a publicly available *full* version of [43] and the following comparison is w.r.t. the current status of a draft of [43] that the authors kindly shared with us.

4. To prove *upper* bounds on capacity, it is also an old idea to extract structured subcodes from *any* infinite sequence of good codes. Depending on the applications, the nature of *structures* and techniques used to extract them may vary. To the best of our knowledge, in coding theory, the use of Ramsey theory for obtaining symmetric subcodes dates back to as least as early as Blinovsky [9]. His techniques were applied in a similar manner in followup works by Polyanskiy [34] and Alon–Bukh–Polyanskiy [2]. The work in [43] generalized this idea and managed to extract structured subcodes from arbitrary codes for *general* adversarial channels. Since they dealt with unique decoding, *pairwise* equicoupledness suffices. In our setup, we would like a sequence of subcodes which are *L-wise equicoupled* in the sense that the (empirical) joint distribution of any L-tuple of codewords from the extracted subcode is approximately the same and is close to some $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$. This resembles but generalizes Polyanskiy's [34] techniques. One of the downsides of invoking Ramsey theory is that the reduction usually causes terrible detriment to the rate of the code, since the smallest size for a combinatorial object to contain abundant structures is generally poorly understood in combinatorics. However, we are fine to tolerate such a rate loss since we only care about the *positivity* of the $(L-1)$-list decoding capacity.
5. To show *lower* bounds on capacity, we use the random coding argument aided by *expurgation*. In the prior work [43], the achievability result is obtained by greedy packing. This is reminiscent of a classic technique in coding theory for proving the existence of good codes of certain size. Since in the unique decoding (hard packing) setting, goodness of a code relies merely on pairwise statistics, the size of a greedy packing can be lower bounded using a standard volume counting argument. Indeed, this idea can be implemented in the general setting by counting the volume of the "forbidden region" of any codeword [43]. However, in list decoding setting, the notion of *confusability* is defined for *tuples* of codewords and translates to bounded multiplicity of intersection of forbidden regions of codewords. It is thus not clear how to pack codewords in a greedy manner while ensuring non-existence of local dense clusters. Instead, our code construction is more information-theoretic. We apply ideas of random coding with expurgation which is commonly used in the study of error exponent in Information Theory. A random code may be mildly locally clustered, but this only occurs at rare locations in the space of all length-n input sequences. Indeed, we are able to show that, with high probability, a random code carefully massaged by shoveling off a small number of codewords attains a GV-type bound for general channels.
6. The most difficult part of our work is the converse.
   a. First assume that the distribution $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ associated to the subcode obtained by Ramsey reduction is *symmetric*. To show that no large $(L-1)$-list decodable code exists for general adversarial channels when $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is not completely positive, we provide upper and lower bounds on the average (over all L-tuples in the equicoupled subcode) inner product between the empirical distribution of an L-tuple and a copositive witness of non-complete positivity of $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$. The bounds contradict each other if the code size exceeds certain constant (independent of the blocklength). We review this *double counting trick* (for unique and list decoding under special settings that appeared in prior work) in Section 5. The $L = 2$ case is proved in [43]. The existence of the witness of non-complete positivity is guaranteed by the duality of certain matrix cones. We generalize calculations in [43] to joint distributions of > 2 random variables. Similar notions of complete positivity and copositivity for tensors exist in the literature and the duality continues to hold.

b. If $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is *asymmetric*, we use a completely different argument. We reduce the claim, in a nontrivial way, to the $L = 2$ case which is known to be true [43]. The $L = 2$ case itself is proved [43] by viewing the task of constructing a long sequence of random variables with prescribed asymmetric pairwise marginals as a zero sum game and using discrete Fourier analysis to provide conflicting bounds on the value of the game, if the sequence is longer than certain constant (again independent of the blocklength).

5 Prior work

Among various ideas, our results are built upon prior work which applies a *double counting trick* to obtain upper bounds on code sizes. We first review this technique which can be found in the proof of the classic Plotkin bound and its generalizations.

5.1 Plotkin [33]

One way to prove Theorem 1 is by lower and upper bounding the expected pairwise distance of any given code $\mathcal{C}$ with minimum distance larger than $2np$ ($p = 1/4 + \epsilon$)

$$\mathop{\mathbb{E}}_{(\underline{x},\underline{x}')\sim\mathcal{C}\times\mathcal{C}}\left[d_{\mathrm{H}}\left(\underline{x},\underline{x}'\right)\right], \tag{2}$$

where $\underline{x}, \underline{x}'$ are uniformly and independently picked from $\mathcal{C}$. First note that pairs $\underline{x} = \underline{x}'$ do not contribute to the expectation. On the one hand, the expectation is clearly at least

$$\frac{|\mathcal{C}|(|\mathcal{C}|-1)}{|\mathcal{C}|^2} d_{\min} > |\mathcal{C}|^{-1}(|\mathcal{C}|-1)2np = |\mathcal{C}|^{-1}(|\mathcal{C}|-1)2n(1/4+\epsilon).$$

On the other hand, if we stack codewords into a $2^{nR} \times n$ matrix and let S_j denote the number of 1's in the j-th column, then from the column's perspective, the above expectation is at most

$$\frac{1}{|\mathcal{C}|^2}\sum_{j=1}^{n} 2S_j(|\mathcal{C}| - S_j).$$

The coefficient 2 is because we need to count $(\underline{x}, \underline{x}')$ and $(\underline{x}', \underline{x})$ separately. This bound is at most $n/2$ by concavity of the summands in S_j. Comparing the upper and lower bounds we have that $|\mathcal{C}| \leqslant 1 + \frac{1}{4\epsilon}$, as claimed in Theorem 1.

5.2 Blinovsky [9]

The above double counting argument can be generalized to the setting of list decoding. For the $(p, L-1)$-list decoding setup we introduced in Definition 2, the earliest work we are aware of following this idea is the one by Blinovsky [9].

Unlike Theorem 1, not only did Blinovsky show that any $(p, L-1)$-list decodable code has to be small as long as $p > P_{L-1}$, he even gave an upper bound (it is still essentially the best as far as we know) on the $(p, L-1)$-list decoding capacity for *any* L. We sketch his idea below but omit the complicated calculations.

First note that proving upper bounds on C_{L-1} for fixed p is equivalent to proving upper bounds on p for fixed rate R. We then define the following three quantities

$$r_{\mathrm{LD}} = \min_{\mathcal{L}\in\binom{\mathcal{C}}{L}} \min_{\underline{y}\in\mathbb{F}_2^n} \max_{\underline{x}\in\mathcal{L}} d_{\mathrm{H}}\left(\underline{y},\underline{x}\right), \tag{3}$$

$$r_{\mathrm{avg}} = \min_{\mathcal{L}\in\binom{\mathcal{C}}{L}} \min_{\underline{y}\in\mathbb{F}_2^n} \mathop{\mathbb{E}}_{\underline{x}\sim\mathcal{L}} \left[d_{\mathrm{H}}\left(\underline{y},\underline{x}\right)\right], \tag{4}$$

$$r_{\mathrm{DC}} = \mathop{\mathbb{E}}_{\mathcal{L}\sim\binom{\mathcal{C}}{L}} \min_{\underline{y}\in\mathbb{F}_2^n} \mathop{\mathbb{E}}_{\underline{x}\sim\mathcal{L}} \left[d_{\mathrm{H}}\left(\underline{y},\underline{x}\right)\right]. \tag{5}$$

All expectations are over uniform selections from the corresponding sets. Namely,

$$\mathop{\mathbb{E}}_{\mathcal{L}\sim\binom{\mathcal{C}}{L}}[\cdot] = \frac{1}{\binom{|\mathcal{C}|}{L}} \sum_{\mathcal{L}\in\binom{\mathcal{C}}{L}} [\cdot], \quad \mathop{\mathbb{E}}_{\underline{x}\sim\mathcal{L}}[\cdot] = \frac{1}{L}\sum_{\underline{x}\in\mathcal{L}}[\cdot].$$

Let us parse what these quantities are measuring.

1. r_{LD} is known as the *list decoding radius* of a given code $\mathcal{C}$. The minimax expression associated to a set $\mathcal{L}$ of vectors

   $$r_{\mathrm{Cheb}} \coloneqq \min_{\underline{y}\in\mathbb{F}_2^n} \max_{\underline{x}\in\mathcal{L}} d_{\mathrm{H}}\left(\underline{y},\underline{x}\right)$$

   is known as the *Chebyshev radius* of $\mathcal{L}$. It is the radius of the smallest circumscribed ball of $\mathcal{L}$. And

   $$p^*(R) \coloneqq \limsup_{n\to\infty} \max_{\mathcal{C}\subset\mathbb{F}_2^n\colon |\mathcal{C}|\geqslant 2^{nR}} r_{\mathrm{LD}}(\mathcal{C})$$

   is precisely the largest allowable p for $(p, L-1)$-list decodable codes of a fixed rate R to exist. Note that $p^*(0) = P_{L-1}$.
2. r_{avg} is known as the *average list decoding radius* and the min-average expression

   $$\min_{\underline{y}\in\mathbb{F}_2^n} \mathop{\mathbb{E}}_{\underline{x}\sim\mathcal{L}} \left[d_{\mathrm{H}}\left(\underline{y},\underline{x}\right)\right]$$

   is the *average radius* of a list. It is not hard to see that the average radius center of $\mathcal{L}$ is the component-wise majority of vectors in $\mathcal{L}$, i.e., the minimizer $\underline{y}^*$ has $\mathsf{MAJ}\left(\underline{x}(i)\colon \underline{x}\in\mathcal{L}\right)$ as its i-th component. Define *plurality* as

   $$\begin{array}{rccl} \mathsf{PLUR}\colon & \mathbb{F}_2^L & \to & [0,1] \\ & (x_1,\cdots,x_L) & \mapsto & \frac{1}{L}\left|\{i\in[L]\colon x_i = \mathsf{MAJ}(x_1,\cdots,x_L)\}\right|, \end{array}$$

   which is the fraction of the most frequent symbol. Then the average radius of $\mathcal{L}$ can be explicitly written as

   $$\min_{\underline{y}\in\mathbb{F}_2^n} \mathop{\mathbb{E}}_{\underline{x}\sim\mathcal{L}} \left[d_{\mathrm{H}}\left(\underline{y},\underline{x}\right)\right] = \sum_{j=1}^{n} \left(1 - \mathsf{PLUR}\left(\underline{x}(j)\colon \underline{x}\in\mathcal{L}\right)\right).$$

3. r_{DC} is a further variant of r_{LD} – the ultimate quantity we are looking for. It is the object that Blinovsky was really dealing with. Note that it is in the same spirit as the quantity (2) considered in the double counting argument in the proof of the classic Plotkin bound. Blinovsky used r_{DC} as a proxy to finally bound r_{LD}.

By extracting a constant weight subcode and applying the double counting trick (and using the convexity of a certain function), Blinovsky showed the following

▶ **Lemma 6.** *Let $\lambda\in[0,1/2]$ and fix $R = 1 - H(\lambda)$. Then*

$$r_{\mathrm{DC}} \leqslant \sum_{i=1}^{\lceil L/2\rceil} \frac{\binom{2i-2}{i-1}}{i} (\lambda(1-\lambda))^i.$$

Apparently, by definition, we have

$$r_{\mathrm{LD}} \geqslant r_{\mathrm{avg}}, \quad r_{\mathrm{DC}} \geqslant r_{\mathrm{avg}}.$$

So Lemma 6 automatically holds for r_{avg}. However, a priori the relation between r_{LD} and r_{DC} is unclear. Surprisingly, Blinovsky showed that it is "okay" to replace the first and third optimization in r_{LD} with averaging in the sense that the following holds.

▶ **Lemma 7.** *For any infinite sequence of codes $\{\mathcal{C}_n\}_n$, there exists an infinite sequence of subcodes $\mathcal{C}'_n \subseteq \mathcal{C}_n$ such that $r_{\mathrm{LD}}(\mathcal{C}') = r_{\mathrm{avg}}(\mathcal{C}') + o(n)$.*

The proof involves an *equidistant* subcode extraction step using Ramsey theory. Lemma 7 implies that the same bound in Lemma 6 holds for r_{LD} as well!

5.3 Cohen–Litsyn–Zémor [15]

Similar ideas were used to provide upper bounds on the erasure list decoding capacity. A binary code is said to be $(p, L-1)$-*erasure list decodable* if for any $\mathcal{T} \in \binom{[n]}{n(1-p)}$ and any $\underline{y} \in \mathbb{F}_2^{(1-p)n}$, $\left|\{\underline{x} \in \mathcal{C} \colon \underline{x}|_{\mathcal{T}} = \underline{y}\}\right| \leqslant L-1$, where $\underline{x}|_{\mathcal{T}}$ denotes the restriction of $\underline{x}$ to $\mathcal{T}$, i.e., a vector of length $|\mathcal{T}|$ only consisting of components from $\underline{x}$ indexed by elements in $\mathcal{T}$. The erasure list decoding radius $r_{\mathrm{LD,eras}}$ and the $(p, L-1)$-erasure list decoding capacity $C_{L-1,\mathrm{eras}}$ are defined in the same manner. Cohen–Litsyn–Zémor [15] showed that

▶ **Theorem 8** ([15])**.** *$C_{L,\mathrm{eras}} \leqslant 1 - H(\lambda)$, where λ is the unique root of the equation $\lambda^{L+1} + (1-\lambda)^{L+1} = 1 - p$ in $[0, 1/2]$.*

The idea is essentially still double counting. Here, it turns out that the right object to be counted is the *erasure radius* of a list $\mathcal{L}$,

$$r_{\mathrm{eras}} := \left|\{i \in [n] \colon \underline{x}(i) \text{ are the same } \forall \underline{x} \in \mathcal{L}\}\right|.$$

Extracting a subcode living on a sphere (followed by shifting out the center to get a constant weight subcode $\mathcal{C}'$) and conducting similar calculations on

$$\mathop{\mathbb{E}}_{\mathcal{L} \sim \binom{\mathcal{C}'}{L}} \left[r_{\mathrm{eras}}(\mathcal{L})\right],$$

allow the authors to conclude Theorem 8.

▶ Remark 9. The original paper [15] was stated for *generalized distance* which is equivalent to erasure list decoding radius via a well-known connection. The above version was presented in Guruswami's PhD thesis [24].

5.4 Wang–Budkuley–Bogdanov–Jaggi [43]

As mentioned, our work is a continuation of the prior work [43] which a subset of the authors were involved in. We refer the readers to the corresponding paragraphs in Sec. 1 and Sec. 3 for a review of their work along with a comparison with this work.

6 Organization of the paper

In Sec. 1 we have seen numeric examples that illustrate our results. In Sec. 2 we properly motivated the problem and introduced relevant background in coding theory. Our contributions in this paper were listed in details in Sec. 3. In Sec. 4 we reviewed various techniques used in this paper and highlighted our innovations. Prior works that our results build up on and push forward were surveyed in Sec. 5.

The rest of the paper is organized as follows. We fix our notational conventions in Sec. 7 and provide necessary preliminaries, especially *the method of types* in Information Theory, in Sec. 8. We develop basic notions that will be used throughout the paper in Sec. 9. In particular, *general adversarial channels* and objects associated to them will be introduced in this section. In Sec. 10 we prove the list decoding capacity theorem for general adversarial channels when L is asymptotically large. Furthermore, we obtain *tight* list size bounds for *most* capacity-achieving constant composition codes in Sec. 11. In Sec. 12 and Sec. 13 we show sufficiency and necessity, respectively, of the criterion for the existence of exponential-sized $(L-1)$-list decodable codes (where L is an arbitrary universal constant) for general adversarial channels. In Sec. 14 we make two remarks on the converse, which is technically the most challenging piece of our work. In Sec. 15 we verify the correctness of our characterization obtained in Sec. 12 and Sec. 13 by running it on the problem specialized to the bit-flip model which has been understood in prior works [9, 2]. In Sec. 16, utilizing tools developed and facts proved in this paper, we rigorously rederive Blinovsky's [9] results. We obtain more intuitive expressions and demystify his calculations. In Sec. 17 we evaluate bounds on the unique decoding capacity ($L = 2$) in [43] under the bit-flip model. We conclude the paper and list several open questions and future directions in Sec. 18. Some calculations and background knowledge are deferred to Appendices A, B, C and D.

7 Notation

Conventions. Sets are denoted by capital letters in calligraphic typeface, e.g., $\mathcal{C}, \mathcal{I}$, etc.. Random variables are denoted by lower case letters in boldface or capital letters in plain typeface, e.g., $\mathbf{m}, \mathbf{x}, \mathbf{s}, U, W$, etc.. Their realizations are denoted by corresponding lower case letters in plain typeface, e.g., m, x, s, u, w, etc.. Vectors (random or fixed) of length n, where n is the blocklength without further specification, are denoted by lower case letters with underlines, e.g., $\underline{\mathbf{x}}, \underline{\mathbf{s}}, \underline{x}, \underline{s}$, etc.. The i-th entry of a vector $\underline{x} \in \mathcal{X}^n$ is denoted by $\underline{x}(i)$ since we can alternatively think of $\underline{x}$ as a function from $[n]$ to $\mathcal{X}$. Same for a random vector $\underline{\mathbf{x}}$. Matrices are denoted by capital letters in boldface, e.g., $\mathbf{P}, \mathbf{\Sigma}$, etc.. Similarly, the (i,j)-th entry of a matrix $\mathbf{G} \in \mathbb{F}^{n\times m}$ is denoted by $\mathbf{G}(i,j)$. We sometimes write $\mathbf{G}_{n\times m}$ to explicitly specify its dimension. For square matrices, we write $\mathbf{G}_n$ for short. Letter $\mathbf{I}$ is reserved for identity matrix. Tensors are denoted by capital letters in plain typeface, e.g., T, P, etc..

Functions. We use the standard Bachmann–Landau (Big-Oh) notation for asymptotics of real-valued functions in positive integers.

For $x \in \mathbb{R}$, let $[x]^+ \coloneqq \max\{x, 0\}$.

For two real-valued functions f, g on the same domain Ω, let fg and f/g denote the functions obtained by multiplying and taking the ratio of the images of f and g point-wise, respectively. That is, for $\omega \in \Omega$,

$$(fg)(\omega) \coloneqq f(\omega)g(\omega), \quad (f/g)(\omega) \coloneqq f(\omega)/g(\omega).$$

In particular, for types or distributions, we can write $\tau_{\mathbf{x},\mathbf{y}} = \tau_{\mathbf{x}}\tau_{\mathbf{y}|\mathbf{x}}, \tau_{\mathbf{y}|\mathbf{x}} = \tau_{\mathbf{x},\mathbf{y}}/\tau_{\mathbf{x}}$, or $P_{\mathbf{x},\mathbf{y}} = P_{\mathbf{x}}P_{\mathbf{y}|\mathbf{x}}, P_{\mathbf{y}|\mathbf{x}} = P_{\mathbf{x},\mathbf{y}}/P_{\mathbf{x}}$ and so on.

For two real-valued functions $f(n), g(n)$ in positive integers, we say that $f(n)$ *asymptotically equals* $g(n)$, denoted $f(n) \asymp g(n)$, if

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1.$$

For instance, $2^{n+\log n} \asymp 2^{n+\log n} + 2^n$, $2^{n+\log n} \not\asymp 2^n$. We write $f(n) \doteq g(n)$ (read $f(n)$ dot equals $g(n)$) if the coefficients of the dominant terms in the exponents of $f(n)$ and $g(n)$ match,

$$\lim_{n\to\infty} \frac{\log f(n)}{\log g(n)} = 1.$$

For instance, $2^{3n} \doteq 2^{3n+n^{1/4}}$, $2^{2^n} \neq 2^{2^{n+\log n}}$. Note that $f(n) \asymp g(n)$ implies $f(n) \doteq g(n)$, but the converse is not true.

For any $q \in \mathbb{R}_{>0}$, we write $\log_q(\cdot)$ for the logarithm to the base q. In particular, let $\log(\cdot)$ and $\ln(\cdot)$ denote logarithms to the base two and e, respectively.

Sets. For any two sets $\mathcal{A}$ and $\mathcal{B}$ with additive and multiplicative structures, let $\mathcal{A}+\mathcal{B}$ and $\mathcal{A}\cdot\mathcal{B}$ denote the Minkowski sum and Minkowski product of them which are defined as

$$\mathcal{A}+\mathcal{B} \coloneqq \{a+b\colon a\in\mathcal{A}, b\in\mathcal{B}\}, \quad \mathcal{A}\cdot\mathcal{B} \coloneqq \{a\cdot b\colon a\in\mathcal{A}, b\in\mathcal{B}\},$$

respectively. If $\mathcal{A}=\{x\}$ is a singleton set, we write $x+\mathcal{B}$ and $x\cdot\mathcal{B}$ for $\{x\}+\mathcal{B}$ and $\{x\}\cdot\mathcal{B}$.

For any finite set $\mathcal{X}$ and any integer $0 \leqslant k \leqslant |\mathcal{X}|$, we use $\binom{\mathcal{X}}{k}$ to denote the collection of all subsets of $\mathcal{X}$ of size k.

$$\binom{\mathcal{X}}{k} \coloneqq \{\mathcal{Y}\subseteq\mathcal{X}\colon\ |\mathcal{Y}|=k\}.$$

For $M\in\mathbb{Z}_{>0}$, we let $[M]$ denote the set of first M positive integers $\{1,2,\cdots,M\}$.

For any $\mathcal{A}\subseteq\Omega$, the indicator function of $\mathcal{A}$ is defined as, for any $x\in\Omega$,

$$\mathbb{1}_{\mathcal{A}}(x) \coloneqq \begin{cases} 1, & x\in\mathcal{A} \\ 0, & x\notin\mathcal{A} \end{cases}.$$

At times, we will slightly abuse notation by saying that $\mathbb{1}_{\mathsf{A}}$ is 1 when event A happens and 0 otherwise. Note that $\mathbb{1}_{\mathcal{A}}(\cdot) = \mathbb{1}_{\{\cdot\in\mathcal{A}\}}$.

Geometry. For any $\underline{x}\in\mathbb{F}_q^n$, let $wt_{\mathrm{H}}(\underline{x})$ denote the Hamming weight of $\underline{x}$, i.e., the number of nonzero entries of $\underline{x}$.

$$wt_{\mathrm{H}}(\underline{x}) \coloneqq |\{i\in[n]\colon \underline{x}(i)\neq 0\}|.$$

For any $\underline{x},\underline{y}\in\mathbb{F}_q^n$, let $d_{\mathrm{H}}(\underline{x},\underline{y})$ denote the Hamming distance between $\underline{x}$ and $\underline{y}$, i.e., the number of locations where they differ.

$$d_{\mathrm{H}}(\underline{x},\underline{y}) \coloneqq wt_{\mathrm{H}}(\underline{x}-\underline{y}) = \left|\left\{i\in[n]\colon \underline{x}(i)\neq\underline{y}(i)\right\}\right|.$$

Balls and spheres in $\mathbb{F}_q^n$ centered around some point $\underline{x}\in\mathbb{F}_q^n$ of certain radius $r\in\{0,1,\cdots,n\}$ w.r.t. the Hamming metric are defined as follows.

$$\mathcal{B}_{\mathrm{H}}^n(\underline{x},r) \coloneqq \left\{\underline{y}\in\mathbb{F}_q^n\colon d_{\mathrm{H}}(\underline{x},\underline{y})\leqslant r\right\}, \quad \mathcal{S}_{\mathrm{H}}^n(\underline{x},r) \coloneqq \left\{\underline{y}\in\mathbb{F}_q^n\colon d_{\mathrm{H}}(\underline{x},\underline{y})= r\right\}.$$

We will drop the subscript and superscript for the associated metric and dimension when they are clear from the context.

Probability. The probability mass function (p.m.f.) of a discrete random variable $\mathbf{x}$ or a random vector $\underline{\mathbf{x}}$ is denoted by $P_{\mathbf{x}}$ or $P_{\underline{\mathbf{x}}}$. Here we use the following shorthand notation to denote the probability that $\mathbf{x}$ or $\underline{\mathbf{x}}$ distributed according to $P_{\mathbf{x}}$ or $P_{\underline{\mathbf{x}}}$ takes a particular value.

$$P_{\mathbf{x}}(x) \coloneqq \Pr_{\mathbf{x}\sim P_{\mathbf{x}}}[\mathbf{x} = x], \quad P_{\underline{\mathbf{x}}}(\underline{x}) = \Pr_{\underline{\mathbf{x}}\sim P_{\underline{\mathbf{x}}}}[\underline{\mathbf{x}} = \underline{x}],$$

for any $x \in \mathcal{X}$ or $\underline{x} \in \mathcal{X}^n$. If every entry of $\underline{\mathbf{x}}$ is independently and identically distributed (i.i.d.) according to $P_{\mathbf{x}}$, then we write $\underline{\mathbf{x}} \sim P_{\mathbf{x}}^{\otimes n}$, where $P_{\mathbf{x}}^{\otimes n}$ is a product distribution defined as

$$P_{\underline{\mathbf{x}}}(\underline{x}) = P_{\mathbf{x}}^{\otimes n}(\underline{x}) \coloneqq \prod_{i=1}^{n} P_{\mathbf{x}}(\underline{x}(i)).$$

For a finite set $\mathcal{X}$, $\Delta(\mathcal{X})$ denotes the probability simplex on $\mathcal{X}$, i.e., the set of all probability distributions supported on $\mathcal{X}$,

$$\Delta(\mathcal{X}) \coloneqq \left\{ P_{\mathbf{x}} \in [0,1]^{|\mathcal{X}|} \colon \sum_{x\in\mathcal{X}} P_{\mathbf{x}}(x) = 1 \right\}.$$

Similarly, $\Delta(\mathcal{X}\times\mathcal{Y})$ denotes the probability simplex on $\mathcal{X}\times\mathcal{Y}$,

$$\Delta(\mathcal{X}\times\mathcal{Y}) \coloneqq \left\{ P_{\mathbf{x},\mathbf{y}} \in [0,1]^{|\mathcal{X}|\times|\mathcal{Y}|} \colon \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_{\mathbf{x},\mathbf{y}}(x,y) = 1 \right\}.$$

Let $\Delta(\mathcal{Y}|\mathcal{X})$ denote the set of all conditional distributions,

$$\Delta(\mathcal{Y}|\mathcal{X}) \coloneqq \left\{ P_{\mathbf{y}|\mathbf{x}} \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{Y}|} \colon P_{\mathbf{y}|\mathbf{x}}(\cdot|x) \in \Delta(\mathcal{Y}),\ \forall x \in \mathcal{X} \right\}.$$

The general notion for multiple spaces is defined in the same manner.

Let $\mathrm{Unif}(\Omega)$ denote the uniform distribution on some probability space Ω.

For a joint distribution $P_{\mathbf{x},\mathbf{y}} \in \Delta(\mathcal{X}\times\mathcal{Y})$, let $[P_{\mathbf{x},\mathbf{y}}]_{\mathbf{x}} \in \Delta(\mathcal{X})$ denote the *marginalization* onto the variable $\mathbf{x}$, i.e., for $x \in \mathcal{X}$,

$$[P_{\mathbf{x},\mathbf{y}}]_{\mathbf{x}}(x) \coloneqq \sum_{y\in\mathcal{Y}} P_{\mathbf{x},\mathbf{y}}(x,y).$$

Sometimes we simply write it as $P_{\mathbf{x}}$ when the notation is not overloaded.

Algebra. Let $\|\cdot\|_p$ denote the standard ℓ^p-norm. Specifically, for any $\underline{x} \in \mathbb{R}^n$,

$$\|\underline{x}\|_p \coloneqq \left(\sum_{i=1}^{n} |\underline{x}(i)|^p \right)^{1/p}.$$

For brevity, we also write $\|\cdot\|$ for the ℓ^2-norm.

An order-k dimension-$(n_1,\cdots,n_k)$ tensor T is a multidimensional array. It can be thought as a function on the product space $[n_1]\times\cdots\times[n_k]$ which identifies the value of each of its entries.

$$\begin{array}{rccl} T\colon & [n_1]\times\cdots\times[n_k] & \to & \mathbb{R} \\ & (i_1,\cdots,i_k) & \mapsto & T(i_1,\cdots,i_k), \end{array}$$

where, as usual, we use $T(i_1,\cdots,i_k)$ to denote its $(i_1,\cdots,i_k)$-th entry.

Without specification, all matrices and tensors are over the real number field. The space of $n \times m$ matrices is denoted by

$$\mathsf{Mat}_{n\times m} := \left\{\mathbf{M} \in \mathbb{R}^{n\times m}\right\} \cong \mathbb{R}^{n\cdot m}.$$

When $n = m$, we write Mat_n for the space of square matrices of dimension n. The space of order-k dimension-$(n_1, \cdots, n_k)$ tensors is denoted by

$$\mathsf{Ten}^{\otimes k}_{n_1,\cdots,n_k} := \left\{T \in \mathbb{R}^{n_1\times\cdots\times n_k}\right\} \cong \mathbb{R}^{n_1\cdots n_k}.$$

If every dimension of T is the same, $n_1 = \cdots = n_k = n$, then we write $\mathsf{Ten}^{\otimes k}_n$ for the space of equilateral tensors of order k and dimension n. Definitions of the sets of *symmetric* (Sym), *non-negative* (NN), *doubly non-negative* (DNN), *positive semidefinite* (PSD), *completely positive* (CP), *copositive* (coP), etc., matrices and tensors are deferred to the corresponding sections where we need them. Note that $\mathsf{Mat}_{n,m} = \mathsf{Ten}^{\otimes 2}_{n,m}$. When the order of the tensors is $k = 2$, namely matrices, we drop the superscript $\otimes 2$.

For a tensor $T \in \mathsf{Ten}^{\otimes k}_{n_1,\cdots,n_k}$, we use $\|T\|_{\mathrm{F}}$ to denote the *Frobenius norm* of T, which is the ℓ^2 norm when T is vectorized into a length-$n_1\cdots n_k$ vector.

$$\|T\|_{\mathrm{F}} := \left(\sum_{(i_1,\cdots,i_k)\in[n_1]\times\cdots\times[n_k]} T(i_1,\cdots,i_k)^2\right)^{1/2}.$$

We use $\|T\|_{\mathrm{sav}}$ to denote the *sum-absolute-value norm* of T which is the ℓ^1 norm after vectorization.

$$\|T\|_{\mathrm{sav}} := \sum_{(i_1,\cdots,i_k)\in[n_1]\times\cdots\times[n_k]} |T(i_1,\cdots,i_k)|.$$

Similarly, define

$$\|T\|_{\mathrm{mav}} := \max_{(i_1,\cdots,i_k)\in[n_1]\times\cdots\times[n_k]} |T(i_1,\cdots,i_k)|$$

to be the *max-absolute-value* norm of T, which is the ℓ^∞ norm when viewed as a vector.

Note that the Frobenius norm, sum-absolute-value norm and max-absolute-value are different from the matrix/tensor 2-norm, 1-norm and ∞-norm. Though they do trivially coincide with the corresponding vector norm when the order of the tensor is one.

We endow the matrix/tensor space with an inner product. For tensors T_1 and T_2 both in $\mathsf{Ten}^{\otimes k}_{n_1,\cdots,n_k}$,

$$\langle T_1, T_2\rangle := \sum_{(i_1,\cdots,i_k)\in[n_1]\times\cdots\times[n_k]} T_1(i_1,\cdots,i_k)T_2(i_1,\cdots,i_k).$$

When T_1, T_2 are matrices, the above definition agrees with the *Frobenius inner product*, which is alternatively defined as $\mathrm{Tr}\left(T_1^\top T_2\right)$. When T_1, T_2 are vectors, this inner product becomes the standard inner product associated to $\mathbb{R}^n$ as a Hilbert space, which is denoted by the same notation without confusion.

Let S_n denote the *symmetric group* of degree n consisting of $n!$ permutations on $[n]$. Permutations are typically denoted by lower case Greek letters.

Information theory. We use $H(\cdot)$ to interchangeably denote the binary entropy function and the Shannon entropy; the exact meaning will be clear from the context. In particular, for any $p \in [0,1]$, $H(p)$ denotes the binary entropy

$$H(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}.$$

For a distribution $P \in \Delta(\mathcal{X})$ on a finite alphabet $\mathcal{X}$ or a random variable $\mathbf{x} \sim P$ distributed according to P, the Shannon entropy of P or $\mathbf{x}$ is defined similarly as

$$H(P) = H(\mathbf{x}) \coloneqq \sum_{x \in \mathcal{X}} P(x) \log \frac{1}{P(x)}.$$

For two distributions $P, Q \in \Delta(\mathcal{X})$ on the same alphabet $\mathcal{X}$, the *Kullback–Leibler (KL) divergence* between them is defined as

$$D(P \| Q) \coloneqq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

If $\mathbf{x}, \mathbf{y}$ are jointly distributed according to $P_{\mathbf{x},\mathbf{y}} \in \Delta(\mathcal{X} \times \mathcal{Y})$, then their *joint entropy* is defined as

$$H(\mathbf{x}, \mathbf{y}) = H(P_{\mathbf{x},\mathbf{y}}) \coloneqq \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{\mathbf{x},\mathbf{y}}(x,y) \log \frac{1}{P_{\mathbf{x},\mathbf{y}}(x,y)};$$

their *mutual information* is defined as

$$\begin{aligned} I(\mathbf{x}; \mathbf{y}) \coloneqq & D\left(P_{\mathbf{x},\mathbf{y}} \| P_{\mathbf{x}} P_{\mathbf{y}}\right) \\ = & \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{\mathbf{x},\mathbf{y}}(x,y) \log \frac{P_{\mathbf{x},\mathbf{y}}(x,y)}{P_{\mathbf{x}}(x) P_{\mathbf{y}}(y)} \\ = & \sum_{y \in \mathcal{Y}} P_{\mathbf{y}}(y) \sum_{x \in \mathcal{X}} P_{\mathbf{x}|\mathbf{y}}(x|y) \log \frac{P_{\mathbf{x}|\mathbf{y}}(x|y)}{P_{\mathbf{x}}(x)}. \end{aligned}$$

If the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ is $P_{\mathbf{y}|\mathbf{x}} \in \Delta(\mathcal{Y}|\mathcal{X})$, then the *conditional entropy* of $\mathbf{y}$ given $\mathbf{x}$ is defined as

$$\begin{aligned} H(\mathbf{y}|\mathbf{x}) \coloneqq & \sum_{x \in \mathcal{X}} P_{\mathbf{x}}(x) H(\mathbf{y}|\mathbf{x} = x) \\ = & \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} P_{\mathbf{x},\mathbf{y}}(x,y) \log \frac{P_{\mathbf{x}}(x)}{P_{\mathbf{x},\mathbf{y}}(x,y)}. \end{aligned}$$

It is easy to check that different definitions above for the same quantities are consisted with each other.

8 Preliminaries

▶ **Lemma 10** (Stirling's approximation). *For any $n \in \mathbb{Z}_{>0}$, $n! \asymp \sqrt{2\pi n}\,(n/e)^n$.*

▶ **Corollary 11** (Asymptotics of multinomials). *For any positive integers $n \geqslant q$ and any q-partition $(n_1, \cdots, n_q)$ of n ($n_1 + \cdots + n_q = n$, $n_i \geqslant 0$ for every i), $\binom{n}{n_1, \cdots, n_q} \doteq 2^{nH(P)}$, where $P \in \Delta([q])$ is an empirical distribution such that for $i \in [q]$, $P(i) = n_i/n$. More precisely, we have $\binom{n}{n_1, \cdots, n_q} \asymp \nu(n)^{-1} 2^{nH(P)}$, where $\nu(n)$ is a polynomial defined as*

$$\nu(n) \coloneqq (2\pi n)^{\frac{q-1}{2}} \left(\prod_{i=1}^{q} P(i) \right)^{\frac{1}{2}}.$$

▶ **Fact 12** (Approximation of binomials). *For any positive integers $n \geqslant k$,*

$$\left(\frac{n}{k}\right)^k \leqslant \binom{n}{k} \leqslant \left(\frac{en}{k}\right)^k, \tag{6}$$

$$(n-k)^k \leqslant (n-k+1)^k \leqslant \binom{n}{k} \leqslant n^k. \tag{7}$$

Without loss of generality, we write $\mathcal{X} = \{1, \cdots, |\mathcal{X}|\}$. For $\underline{x} \in \mathcal{X}^n$ and $x \in \mathcal{X}$, let

$$N_x(\underline{x}) := \left|\{i \in [n] \colon \underline{x}(i) = x\}\right|,$$

which counts the number of occurrences of a symbol x in a vector $\underline{x}$. Similarly, define

$$N_{x,y}\left(\underline{x}, \underline{y}\right) := \left|\left\{i \in [n] \colon \underline{x}(i) = x,\ \underline{y}(i) = y\right\}\right|.$$

▶ **Definition 13** (Types). *For a length-n vector $\underline{x}$ over a finite alphabet $\mathcal{X}$, the type $\tau_{\underline{x}}$ of $\underline{x}$ is a length-$|\mathcal{X}|$ (empirical) probability vector (or the histogram of $\underline{x}$), i.e., $\tau_{\underline{x}} \in [0,1]^{|\mathcal{X}|}$ has entries $\tau_{\underline{x}}(x) := N_x(\underline{x})/n$ for all $x \in \mathcal{X}$.*

▶ **Definition 14** (Joint types and conditional types). *The* joint type $\tau_{\underline{x},\underline{y}} \in [0,1]^{|\mathcal{X}|\times|\mathcal{Y}|}$ *of two vectors $\underline{x} \in \mathcal{X}^n$ and $\underline{y} \in \mathcal{Y}^n$ is defined as $\tau_{\underline{x},\underline{y}}(x,y) = N_{x,y}(\underline{x},\underline{y})/n$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$.*

The conditional type $\tau_{\underline{y}|\underline{x}} \in [0,1]^{|\mathcal{X}|\times|\mathcal{Y}|}$ of a vector $\underline{y} \in \mathcal{Y}^n$ given another vector $\underline{x} \in \mathcal{X}^n$ is defined as $\tau_{\underline{y}|\underline{x}}(y|x) = N_{x,y}\left(\underline{x},\underline{y}\right)/N_x\left(\underline{x}\right)$.

▶ Remark 15 (Types vs. distributions). Types are empirical distributions of length-n vectors. They can only take rational values, in particular, a/n for $a \in \{0,1,\cdots,n\}$. For finite alphabets and a fixed n, there are only poly(n) many types. However, there are uncountably infinitely many distributions on any finite alphabets and they form a probability simplex.

▶ Remark 16. We will also write $\tau_{\mathbf{x}}, \tau_{\mathbf{x},\mathbf{y}}, \tau_{\mathbf{y}|\underline{x}}, \tau_{\mathbf{y}|\mathbf{x}}$, etc., for generic types that are taken from the corresponding sets of types even if they do not come from instantiated vectors. For instance, $\tau_{\mathbf{x}}$ is a type in $\mathcal{P}^{(n)}(\mathcal{X})$ corresponding to any $\underline{x} \in \mathcal{T}_{\underline{\mathbf{x}}}(\tau_{\mathbf{x}})$. The particular choice of $\underline{x}$ is not important and will not be specified. These notations are for explicitly distinguishing types from distributions.

▶ **Definition 17** (Set of types). *We use $\mathcal{P}^{(n)}(\mathcal{X})$ to denote the set of types of all length-n vectors over $\mathcal{X}$.*

$$\mathcal{P}^{(n)}(\mathcal{X}) = \left\{\tau_{\underline{x}} \colon \underline{x} \in \mathcal{X}^n\right\}.$$

Similarly, define

$$\mathcal{P}^{(n)}(\mathcal{X},\mathcal{Y}) = \left\{\tau_{\underline{x},\underline{y}} \colon \underline{x} \in \mathcal{X}^n,\ \underline{y} \in \mathcal{Y}^n\right\},$$
$$\mathcal{P}^{(n)}(\mathcal{Y}|\underline{x}) = \left\{\tau_{\underline{y}|\underline{x}} \colon \underline{y} \in \mathcal{Y}^n\right\},$$
$$\mathcal{P}^{(n)}(\mathcal{Y}|\mathcal{X}) = \left\{\tau_{\underline{y}|\underline{x}} \colon \underline{x} \in \mathcal{X}^n,\ \underline{y} \in \mathcal{Y}^n\right\}$$

to be

1. *the set of all joint types;*
2. *the set of all conditional types of $\underline{y}$ given a particular $\underline{x}$;*
3. *the set of all conditional types of $\underline{y}$ given some $\underline{x}$,*

respectively.

▶ **Lemma 18** (Types are dense in distributions). *The union of the sets of types of all possible blocklengths is dense in the set of distributions, i.e.,*

$$\bigcup_{n=1}^{\infty} \mathcal{P}^{(n)}(\mathcal{X})$$

is dense in $\Delta(\mathcal{X})$. This holds true for joint types and conditional types as well.

▶ **Lemma 19** (Number of types). *When alphabet sizes are constants, the number of types of length-n vectors is polynomial in n. To be precise, the number of types of length-n vectors over $\mathcal{X}$ is*

$$\left|\mathcal{P}^{(n)}(\mathcal{X})\right| = \binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1}. \tag{8}$$

For a vector $\underline{x} \in \mathcal{X}^n$ of type $\tau_{\underline{x}}$, the number of conditional types of length-n vectors over $\mathcal{Y}$ given $\underline{x}$ is

$$\left|\mathcal{P}^{(n)}(\mathcal{Y}|\underline{x})\right| = \prod_{x\in\mathcal{X}} \binom{\tau_{\underline{x}}(x)n+|\mathcal{Y}|-1}{|\mathcal{Y}|-1}. \tag{9}$$

The number of conditional types of $\mathcal{Y}$-valued vectors given some $\mathcal{X}$-valued vector is

$$\mathcal{P}^{(n)}(\mathcal{Y}|\mathcal{X}) = \sum_{\tau_{\mathbf{x}}\in\mathcal{P}^{(n)}(\mathcal{X})} \prod_{x\in\mathcal{X}} \binom{\tau_{\mathbf{x}}(x)n+|\mathcal{Y}|-1}{|\mathcal{Y}|-1}. \tag{10}$$

The following elementary bounds from [16] are sufficient for the purposes of this paper.

$$\left|\mathcal{P}^{(n)}(\mathcal{X})\right| \leqslant (n+1)^{|\mathcal{X}|},$$
$$\left|\mathcal{P}^{(n)}(\mathcal{Y}|\underline{x})\right| \leqslant \left|\mathcal{P}^{(n)}(\mathcal{Y}|\mathcal{X})\right| \leqslant (n+1)^{|\mathcal{X}|\cdot|\mathcal{Y}|}.$$

▶ **Definition 20** (Type classes). *Define type class $\mathcal{T}_{\underline{\mathbf{x}}}(\tau_{\mathbf{x}})$ w.r.t. a type $\tau_{\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{X})$ as*

$$\mathcal{T}_{\underline{\mathbf{x}}}(\tau_{\mathbf{x}}) \coloneqq \left\{\underline{x} \in \mathcal{X}^n \colon \tau_{\underline{x}} = \tau_{\mathbf{x}}\right\}.$$

Similarly, the joint type class $\mathcal{T}_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\tau_{\mathbf{x},\mathbf{y}})$ w.r.t. a joint type $\tau_{\mathbf{x},\mathbf{y}} \in \mathcal{P}^{(n)}(\mathcal{X}\times\mathcal{Y})$ is defined as

$$\mathcal{T}_{\underline{\mathbf{x}},\underline{\mathbf{y}}}(\tau_{\mathbf{x},\mathbf{y}}) \coloneqq \left\{(\underline{x},\underline{y}) \in \mathcal{X}^n \times \mathcal{Y}^n \colon \tau_{\underline{x},\underline{y}} = \tau_{\mathbf{x},\mathbf{y}}\right\}.$$

The conditional type class $\mathcal{T}_{\underline{\mathbf{y}}|\underline{x}}(\tau_{\mathbf{y}|\underline{x}})$ w.r.t. a conditional type $\tau_{\mathbf{y}|\underline{x}} \in \mathcal{P}^{(n)}(\mathcal{Y}|\underline{x})$ given a vector $\underline{x} \in \mathcal{X}^n$ is defined as

$$\mathcal{T}_{\underline{\mathbf{y}}|\underline{x}}(\tau_{\mathbf{y}|\underline{x}}) \coloneqq \left\{\underline{y} \in \mathcal{Y}^n \colon \tau_{\underline{y}|\underline{x}} = \tau_{\mathbf{y}|\underline{x}}\right\}.$$

The conditional type class $\mathcal{T}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\tau_{\mathbf{y}|\mathbf{x}})$ w.r.t. a conditional type $\tau_{\mathbf{y}|\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{Y}|\mathcal{X})$ is defined as

$$\mathcal{T}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}(\tau_{\mathbf{y}|\mathbf{x}}) \coloneqq \bigcup_{\tau_{\mathbf{x}}\in\mathcal{P}^{(n)}(\mathcal{X})} \mathcal{T}_{\underline{\mathbf{y}}|\underline{x}'}(\tau_{\mathbf{y}|\mathbf{x}}) \tag{11}$$

$$= \left\{\underline{y} \in \mathcal{Y}^n \colon \exists \underline{x}' \in \mathcal{X}^n,\ \tau_{\underline{y}|\underline{x}'} = \tau_{\mathbf{y}|\mathbf{x}}\right\}, \tag{12}$$

where in Eqn. (11) $\underline{x}'$ can be chosen arbitrarily from $\mathcal{T}_{\underline{\mathbf{x}}}(\tau_{\mathbf{x}})$.

▶ **Lemma 21** (Size of type classes)**.**

1. *For any type* $\tau_{\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{X})$, $\left|\mathcal{T}_{\underline{\mathbf{x}}}(\tau_{\mathbf{x}})\right| \doteq 2^{nH(P_{\mathbf{x}})}$.
2. *For any vector* $\underline{x} \in \mathcal{X}^n$ *and any conditional type* $\tau_{\mathbf{y}|\underline{x}} \in \mathcal{P}^{(n)}(\mathcal{Y}|\underline{x})$, $\left|\mathcal{T}_{\underline{\mathbf{y}}|\underline{x}}\left(\tau_{\mathbf{y}|\underline{x}}\right)\right| \doteq 2^{nH(\mathbf{y}|\mathbf{x})}$, *where the conditional entropy is evaluated w.r.t. the joint type* $\tau_{\underline{x}}\tau_{\mathbf{y}|\underline{x}}$.
3. *For any conditional type* $\tau_{\mathbf{y}|\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{Y}|\mathcal{X})$,

$$\left|\mathcal{T}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}\left(\tau_{\mathbf{y}|\mathbf{x}}\right)\right| \doteq 2^{n \max_{\tau_{\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{X})} H(\mathbf{y}|\mathbf{x})},$$

where the conditional entropy is evaluated w.r.t. the joint type $\tau_{\mathbf{x}}\tau_{\mathbf{y}|\mathbf{x}}$.

Proof.

1. The number of sequences $\underline{x} \in \mathcal{X}^n$ of type $\tau_{\mathbf{x}}$ is precisely

$$\binom{n}{n\tau_{\mathbf{x}}(1), \cdots, n\tau_{\mathbf{x}}(|\mathcal{X}|)}$$

and the claim follows from Lemma 10.

2. Given $\underline{x} \in \mathcal{X}^n$, the number of sequences $\underline{y} \in \mathcal{Y}^n$ of conditional type $\tau_{\mathbf{y}|\underline{x}}$ is precisely

$$\prod_{x \in \mathcal{X}} \binom{n\tau_{\underline{x}}(x)}{n\tau_{\mathbf{y}|\underline{x}}(1|x), \cdots, n\tau_{\mathbf{y}|\underline{x}}(|\mathcal{Y}|\,|x)},$$

and the lemma follows from 10.

3. Note that

$$\mathcal{T}_{\underline{\mathbf{y}}|\underline{x}^*}\left(\tau_{\mathbf{y}|\underline{x}^*}\right) \leqslant \left|\mathcal{T}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}\left(\tau_{\mathbf{y}|\mathbf{x}}\right)\right| \leqslant \left|\mathcal{P}^{(n)}(\mathcal{X})\right| \mathcal{T}_{\underline{\mathbf{y}}|\underline{x}^*}\left(\tau_{\mathbf{y}|\underline{x}^*}\right),$$

where $\underline{x}^*$ is chosen arbitrarily from $\mathcal{T}_{\underline{\mathbf{x}}}\left(\tau^*_{\mathbf{x}}\right)$ and[11]

$$\tau^*_{\mathbf{x}} = \operatorname*{argmax}_{\tau_{\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{X})} \left|\mathcal{T}_{\underline{\mathbf{y}}|\underline{x}}\left(\tau_{\underline{\mathbf{y}}|\underline{x}}\right)\right|.$$

The claim follows from Eqn. (8) and the previous claim.

◀

▶ **Lemma 22.** *If* $\underline{\mathbf{x}}$ *is generated using the product distribution* $P_{\mathbf{x}}^{\otimes n}$*, then for any* $\underline{x} \in \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}})$,

$$\Pr\left[\underline{\mathbf{x}} = \underline{x}\right] = 2^{-nH(P_{\mathbf{x}})}.$$

Moreover,

$$\Pr\left[\underline{\mathbf{x}} \in \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}})\right] \asymp \nu(n)^{-1}.$$

Proof. Both claims follow from elementary calculations. For the first one,

$$\begin{aligned}\Pr\left[\underline{\mathbf{x}} = \underline{x}\right] &= \prod_{x \in \mathcal{X}} P_{\mathbf{x}}(x)^{N_x(\underline{x})} \\ &= 2^{\sum_{x \in \mathcal{X}} N_x(\underline{x}) \log P_{\mathbf{x}}(x)} \\ &= 2^{n \sum_{x \in \mathcal{X}} P_{\mathbf{x}}(x) \log P_{\mathbf{x}}(x)} \qquad (13) \\ &= 2^{-nH(P_{\mathbf{x}})},\end{aligned}$$

where Eqn. (13) is because $\tau_{\underline{x}} = P_{\mathbf{x}}$ and hence $N_x(\underline{x})/n = P_{\mathbf{x}}(x)$ for any $x \in \mathcal{X}$.

[11] In the argmax, $\underline{x} \in \mathcal{T}_{\underline{\mathbf{x}}}\left(\tau_{\mathbf{x}}\right)$ is arbitrary as well.

For the second one,

$$\begin{aligned}\Pr\left[\underline{\mathbf{x}} \in \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}})\right] =& \Pr\left[\tau_{\underline{\mathbf{x}}} = P_{\mathbf{x}}\right] \\ =& \binom{n}{nP_{\mathbf{x}}(1), \cdots, nP_{\mathbf{x}}(|\mathcal{X}|)} \prod_{x\in\mathcal{X}} P_{\mathbf{x}}(x)^{nP_{\mathbf{x}}(x)} \\ \asymp& \nu(n)^{-1} 2^{nH(P)} 2^{-nH(P)} \qquad (14) \\ =& \nu(n)^{-1},\end{aligned}$$

where Eqn. (14) is by Corollary 11. ◀

▶ **Lemma 23** (Markov). *For any non-negative random variable X and any positive number x,*

$$\Pr[X \geqslant x] \leqslant \frac{\mathbb{E}[X]}{x}.$$

▶ **Lemma 24** (Chernoff). *Let $X_1, \cdots, X_n$ be independent (not necessarily identically distributed) $\{0,1\}$-valued random variables. Let*

$$X := \sum_{i=1}^{n} X_i.$$

Then

$$\begin{aligned}\Pr[X \geqslant (1+\epsilon)\mathbb{E}[X]] \leqslant& e^{-\frac{\epsilon^2}{3}\mathbb{E}[X]}, \\ \Pr[X \leqslant (1-\epsilon)\mathbb{E}[X]] \leqslant& e^{-\frac{\epsilon^2}{2}\mathbb{E}[X]}, \\ \Pr[X \notin (1\pm\epsilon)\mathbb{E}[X]] \leqslant& 2e^{-\frac{\epsilon^2}{3}\mathbb{E}[X]}.\end{aligned}$$

▶ **Lemma 25** (Sanov). *Let $\mathcal{Q} \subset \Delta(\mathcal{X})$ be a subset of distributions such that it is equal to the closure of its interior. Let $\underline{\mathbf{x}} \sim P_{\mathbf{x}}^{\otimes n}$ be a random vector whose components are i.i.d. according to $P_{\mathbf{x}}$. Clearly $\underline{\mathbf{x}}$ is expected to have type $\mathbb{E}\left[\tau_{\underline{\mathbf{x}}}\right] = P_{\mathbf{x}}$. Sanov's theorem determines the first-order exponent of the probability that the vector empirically looks like coming from some distribution $Q \in \mathcal{Q}$.*

$$\Pr\left[\tau_{\underline{\mathbf{x}}} \in \mathcal{Q}\right] \doteq 2^{-n \inf_{Q\in\mathcal{Q}} D(Q\|P_{\mathbf{x}})}.$$

▶ Remark 26. One can view Sanov's theorem as a particular form of the Chernoff bound. Since $\underline{\mathbf{x}}(i)$'s are independent, it gives the *correct* exponent of $\Pr\left[\tau_{\underline{\mathbf{x}}} \in \mathcal{Q}\right]$ (up to lower order terms) rather than being merely a bound.

▶ **Lemma 27** (Anti-concentration). *Let X be a non-negative random variable. Then*

$$\Pr[X = 0] \leqslant \frac{Var[X]}{\mathbb{E}[X]^2}.$$

▶ **Fact 28** (Binomial identities). *For any non-negative integers n, K and $0 \leqslant k \leqslant n$, we have*

$$\binom{n}{k} = \binom{n}{n-k}, \qquad (15)$$

$$\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}, \qquad (16)$$

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}, \tag{17}$$

$$2^K = \sum_{i=0}^{K} \binom{n}{i}. \tag{18}$$

We list several basic (in)equalities concerning information measures that we will frequently refer to.

▶ **Fact 29** (Information (in)equalities). *The following inequalities hold for any random variables/distributions over finite sets.*

$$\begin{aligned} H(\mathbf{x},\mathbf{y}) =& H(\mathbf{x}) + H(\mathbf{y}|\mathbf{x}) \\ =& H(\mathbf{y}) + H(\mathbf{x}|\mathbf{y}) \\ =& H(\mathbf{x}|\mathbf{y}) + H(\mathbf{y}|\mathbf{x}) + I(\mathbf{x};\mathbf{y}) \\ =& H(\mathbf{x}) + H(\mathbf{y}) - I(\mathbf{x};\mathbf{y}), \\ I(\mathbf{x};\mathbf{y}) =& H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) \\ =& H(\mathbf{y}) - H(\mathbf{y}|\mathbf{x}) \\ =& D\left(P_{\mathbf{x},\mathbf{y}} \| P_{\mathbf{x}} P_{\mathbf{y}}\right). \end{aligned}$$

9 Basic definitions

▶ **Definition 30** (Adversarial channels). *An adversarial channel* $\mathcal{A} = (\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ *(Fig. 3) is a sextuple consisting of*

1. *an input alphabet* $\mathcal{X}$*;*
2. *a set of input constraints* $\lambda_{\mathbf{x}} \subseteq \mathcal{P}^{(n)}(\mathcal{X})$*;*
3. *a noise alphabet* $\mathcal{S}$*;*
4. *a set of noise constraints* $\lambda_{\mathbf{s}} \subseteq \mathcal{P}^{(n)}(\mathcal{S})$*;*
5. *an output alphabet* $\mathcal{Y}$*;*
6. *a channel law given by a transition probability* $W_{\mathbf{y}|\mathbf{x},\mathbf{s}} \in \Delta(\mathcal{Y}|\mathcal{X} \times \mathcal{S})$.

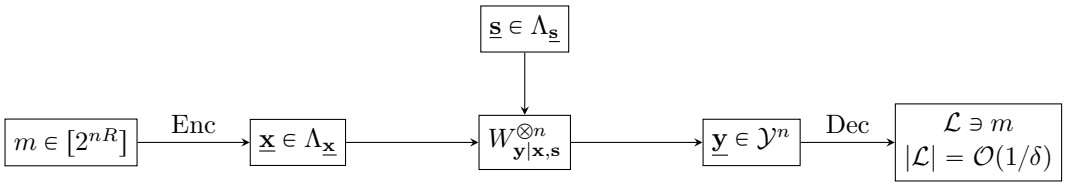


Figure 3 General adversarial channels.

▶ Remark 31. In this paper, we are only concerned with finite alphabets of constant size independent of the blocklength n.

Specifically,

- Though the alphabets $\mathcal{X}, \mathcal{S}$ and $\mathcal{Y}$ can be arbitrary finite sets, it is without loss of generality to realize them using the first $|\mathcal{X}|, |\mathcal{S}|$ and $|\mathcal{Y}|$ positive integers, i.e., $\mathcal{X} = [|\mathcal{X}|], \mathcal{S} = [|\mathcal{S}|]$ and $\mathcal{Y} = [|\mathcal{Y}|]$.[12]

[12] Under such realizations, these sets are *not* necessarily equipped with real arithmetic or modular arithmetic. The metric, if one cares, would be specified by the channel function.

- The input and noise constraint sets $\lambda_{\mathbf{x}}$ and $\lambda_{\mathbf{s}}$ are subsets of types $\mathcal{P}^{(n)}(\mathcal{X})$ and $\mathcal{P}^{(n)}(\mathcal{S})$ respectively. In this paper we assume they are *convex* sets. Since there are polynomially many types in total, we can also think of these collections of types as defined by intersections of hyperplanes or halfspaces, that is, types satisfying a certain finite number of linear (in the entries of the types) (in)equality constraints.
- In this paper, for technical simplicity, we assume that the channel transition function has only *singleton* mass. That is, for each $x \in \mathcal{X}, s \in \mathcal{S}$, $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) = 1$ only for one $y \in \mathcal{Y}$ and is zero for all other outputs. Equivalently, such degenerate distributions can be alternatively thought of as *deterministic* functions

$$\begin{array}{rccc} W\colon & \mathcal{X}\times\mathcal{S} & \to & \mathcal{Y} \\ & (x,s) & \mapsto & y, \end{array}$$

  where y is the *unique* output which is assigned the full probability, $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}(y|x,s) = 1$. Here we slightly abuse the notation and use the same letter for the channel transition distribution and the channel transition function (when the distribution is degenerate). Moreover, we use $\underline{y} = W(\underline{x},\underline{s})$ (with the superscript $\otimes n$ being dropped) to denote the output of n uses of the channel, or equivalently, the n-letter output of the function which acts on $(\underline{x},\underline{s})$ component by component.
  It seems this is a severe restriction (and turns out indeed to be so). Nevertheless, it is still a very first and significant step towards understanding general adversarial channels in full generality. The case where $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ is an arbitrary conditional distribution, or equivalently, the function W is *non-deterministic*, is interesting as well and is left as a future direction.
- For notational convenience, let

$$\begin{aligned} \Lambda_{\underline{\mathbf{x}}} &:= \left\{\underline{x} \in \mathcal{X}^n \colon \tau_{\underline{x}} \in \lambda_{\mathbf{x}}\right\} \\ &= \bigcup_{\tau_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \mathcal{T}_{\underline{\mathbf{x}}}(\tau_{\mathbf{x}}), \\ \Lambda_{\underline{\mathbf{s}}} &:= \left\{\underline{s} \in \mathcal{S}^n \colon \tau_{\underline{s}} \in \lambda_{\mathbf{s}}\right\} \\ &= \bigcup_{\tau_{\mathbf{s}} \in \lambda_{\mathbf{s}}} \mathcal{T}_{\underline{\mathbf{s}}}(\tau_{\mathbf{s}}), \end{aligned}$$

  be sets of codewords and error patterns of admissible types.

▶ **Example 32.** Our framework covers a large family of channel models, including most of the popular and well-studied ones.

1. The standard bit-flip channels. $\mathcal{X} = \mathbb{F}_2, \lambda_{\mathbf{x}} = \mathcal{P}^{(n)}(\mathbb{F}_2), \mathcal{S} = \mathbb{F}_2$,

$$\lambda_{\mathbf{s}} = \left\{\tau_{\mathbf{s}} \in \mathcal{P}^{(n)}(\mathbb{F}_2) \colon \tau_{\mathbf{s}}(1) \leqslant p\right\},$$

   $\mathcal{Y} = \mathbb{F}_2, y = W(x,s) = x \,\mathsf{XOR}\, s$.
2. The standard q-ary channels. $\mathcal{X} = \mathbb{Z}_q$, $\lambda_{\mathbf{x}} = \mathcal{P}^{(n)}(\mathbb{Z}_q)$, $\mathcal{S} = \mathbb{Z}_q$,

$$\lambda_{\mathbf{s}} = \left\{\tau_{\mathbf{s}} \in \mathcal{P}^{(n)}(\mathbb{Z}_q) \colon \tau_{\mathbf{s}}(1) + \cdots + \tau_{\mathbf{s}}(q-1) \leqslant p\right\},$$

   $\mathcal{Y} = \mathbb{Z}_q, y = W(x,s) = x + s \mod q$.
3. The standard erasure channels. $\mathcal{S} = \mathbb{Z}_q, \lambda_{\mathbf{x}} = \mathcal{P}^{(n)}(\mathbb{Z}_q), \mathcal{S} = \mathbb{F}_2$,

$$\lambda_{\mathbf{s}} = \left\{\tau_{\mathbf{s}} \in \mathcal{P}^{(n)}(\mathbb{F}_2) \colon \tau_{\mathbf{s}}(1) \leqslant p\right\},$$

$\mathcal{Y} = \mathbb{Z}_q \cup \{\mathsf{erasure}\}$,

$$y = W(x,s) = \begin{cases} x, & s = 0 \\ \mathsf{erasure}, & s = 1 \end{cases}.$$

4. Weight constrained channels. Any of the above channels with

$$\lambda_{\mathbf{x}} = \left\{ \tau_{\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{X}) \colon 1 - \tau_{\mathbf{x}}(0) \leqslant w \right\}.$$

5. Z-channels (or multiplier/AND channels). $\mathcal{X} = \mathbb{F}_2, \lambda_{\mathbf{x}} = \mathcal{P}^{(n)}(\mathbb{F}_2), \mathcal{S} = \mathbb{F}_2$,

$$\lambda_{\mathbf{s}} = \left\{ \tau_{\mathbf{s}} \in \mathcal{P}^{(n)}(\mathbb{F}_2) \colon \tau_{\mathbf{s}}(1) \leqslant p \right\},$$

$\mathcal{Y} = \mathbb{F}_2$,

$$y = W(x,s) = \begin{cases} 0, & s = 0 \text{ or } x = 0 \\ x, & s = 1 \text{ and } x = 1 \end{cases},$$

or equivalently $y = W(x,s) = x \,\mathsf{AND}\, s$.

6. Adder channels. $\mathcal{X} = \{0, 1, \cdots, q-1\}, \lambda_{\mathbf{x}} = \mathcal{P}^{(n)}(\mathcal{X}), \mathcal{S} = \{0, 1, \cdots, q-1\}$,

$$\lambda_{\mathbf{s}} = \left\{ \tau_{\mathbf{s}} \in \mathcal{P}^{(n)}(\mathcal{S}) \colon \tau_{\mathbf{s}}(1) + \cdots + \tau_{\mathbf{s}}(q-1) \leqslant p \right\},$$

$\mathcal{Y} = \{0, 1, \cdots, 2(q-1)\}, y = W(x,s) = x + s$, where the addition is over $\mathbb{R}$.

7. Noisy typewriter channels. $\mathcal{X} = \mathbb{Z}_q, \lambda_{\mathbf{x}} = \mathcal{P}^{(n)}(\mathbb{Z}_q), \mathcal{S} = \mathbb{F}_2, \lambda_{\mathbf{s}} = \mathcal{P}^{(n)}(\mathbb{F}_2), \mathcal{Y} = \mathbb{Z}_q, y = W(x,s) = x + s \mod q$.
8. OR channels (or $\mathfrak{T}$-channels). $\mathcal{X} = \mathbb{F}_2, \lambda_{\mathbf{x}} = \mathcal{P}^{(n)}(\mathbb{F}_2), \mathcal{S} = \mathbb{F}_2$,

$$\lambda_{\mathbf{s}} = \left\{ \tau_{\mathbf{s}} \in \mathcal{P}^{(n)}(\mathbb{F}_2) \colon \tau_{\mathbf{s}}(1) \leqslant p \right\},$$

$\mathcal{Y} = \mathbb{F}_2, y = W(x,s) = x \,\mathsf{OR}\, s$,

9. Channels under Lee distance. $\mathcal{X} = \mathbb{Z}_q, \lambda_{\mathbf{x}} = \mathcal{P}^{(n)}(\mathbb{Z}_q)$, $\mathcal{S} = \{-\lfloor \frac{q}{2} \rfloor, -\lfloor \frac{q}{2} \rfloor + 1, \cdots, \lfloor \frac{q}{2} \rfloor - 1, \lfloor \frac{q}{2} \rfloor\}$,

$$\lambda_{\mathbf{s}} = \left\{ \tau_{\mathbf{s}} \in \mathcal{P}^{(n)}(\mathcal{S}) \colon \sum_{s=1}^{\lfloor q/2 \rfloor} (\tau_{\mathbf{s}}(s) - \tau_{\mathbf{s}}(-s)) \cdot s \leqslant p \right\},$$

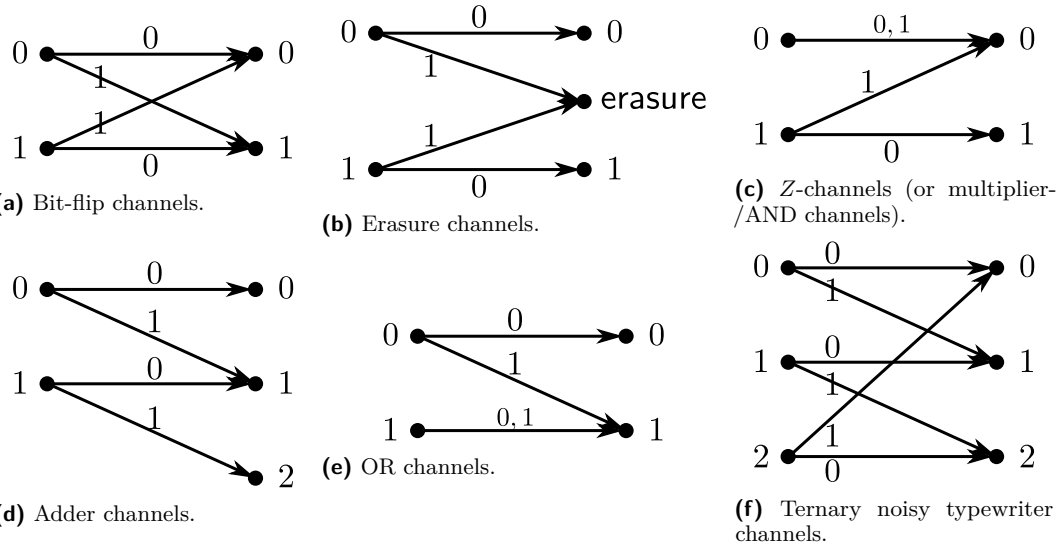
$\mathcal{Y} = \mathbb{Z}_q, y = W(x,s) = x + s$ over the reals.

10. Other more complicated channels, e.g., the one we defined in Sec. 1.

▶ **Definition 33** (Self-couplings). *A joint distribution $P_{\mathbf{x}_1, \cdots, \mathbf{x}_L} \in \Delta(\mathcal{X}^L)$ is said to be a* $(P_{\mathbf{x}}, L)$-self-coupling *for some $P_{\mathbf{x}} \in \Delta(\mathcal{X})$ if all of its marginals equal $P_{\mathbf{x}}$, i.e., $\left[P_{\mathbf{x}_1, \cdots, \mathbf{x}_L}\right]_{\mathbf{x}_i} = P_{\mathbf{x}}$ for all $i \in [L]$. The set of all $(P_{\mathbf{x}}, L)$-self-couplings is denoted by $\mathcal{J}^{\otimes L}(P_{\mathbf{x}})$.*

▶ **Definition 34** (Codes). *In general, a code $\mathcal{C}$ is a subset of $\mathcal{X}^n$. A code $\mathcal{C}$ for an adversarial channel $\mathcal{A} = (\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$ is a subset of $\Lambda_{\underline{\mathbf{x}}}$; n is called the blocklength. Elements in $\mathcal{C}$ are called codewords. The rate $R(\mathcal{C})$ of $\mathcal{C}$ is defined as $R(\mathcal{C}) := (\log |\mathcal{C}|)/n$.*

▶ **Definition 35** (Constant composition codes). *A code $\mathcal{C} \subset \mathcal{X}^n$ is said to be $P_{\mathbf{x}}$-constant composition for some $P_{\mathbf{x}} \in \Delta(\mathcal{X})$ if the type of each codeword is $P_{\mathbf{x}}$, i.e., $\tau_{\underline{x}} = P_{\mathbf{x}}$ for every $\underline{x} \in \mathcal{C}$.*



(a) Bit-flip channels.

(b) Erasure channels.

(c) Z-channels (or multiplier-/AND channels).

(d) Adder channels.

(e) OR channels.

(f) Ternary noisy typewriter channels.

Figure 4 Examples of various well-studied channel models.

▶ **Lemma 36.** *For any code $\mathcal{C} \subset \mathcal{X}^n$ of rate R, there is a constant composition subcode $\mathcal{C}' \subseteq \mathcal{C}$ of asymptotically the same rate.*

Proof. Let $\mathcal{C}' = \mathcal{C} \cap \mathcal{T}_{\underline{\mathbf{x}}}(\tau^*_{\mathbf{x}})$, where

$$\tau^*_{\mathbf{x}} = \operatorname*{argmax}_{\tau_{\mathbf{x}} \in \mathcal{P}^{(n)}(\mathcal{X})} \left|\mathcal{C} \cap \mathcal{T}_{\underline{\mathbf{x}}}(\tau_{\mathbf{x}})\right|$$

is the most common type in $\mathcal{C}$. By Lemma 8 and Lemma 23,

$$\left|\mathcal{C}'\right| \geqslant \frac{|\mathcal{C}|}{(n+1)^{|\mathcal{X}|}} = 2^{nR + |\mathcal{X}| \log(n+1)},$$

which implies that $R(\mathcal{C}') \asymp R(\mathcal{C})$ as n grows. ◀

▶ **Definition 37** (Confusability of tuples of vectors)**.** *A list of L distinct codewords $\underline{x}_1, \cdots, \underline{x}_L \in \mathcal{X}^n$ is said to be* L-confusable *if there are $\underline{y} \in \mathcal{Y}^n$ and $\underline{s}_1, \cdots, \underline{s}_L \in \Lambda_{\underline{\mathbf{s}}}$ such that $W(\underline{x}_i, \underline{s}_i) = \underline{y}$ for all $i \in [L]$.*

▶ **Definition 38** (Confusability of joint distributions)**.** *A $(P_{\mathbf{x}}, L)$-self-coupling $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{J}^{\otimes L}(P_{\mathbf{x}})$ is said to be* L-confusable *if it has some extension given by $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L,\mathbf{s}_1,\cdots,\mathbf{s}_L,\mathbf{y}} \in \Delta\left(\mathcal{X}^L \times \mathcal{S}^L \times \mathcal{Y}\right)$ such that*

1. $[P_{\mathbf{x},\cdots,\mathbf{x}_L,\mathbf{s}_1,\cdots,\mathbf{s}_L,\mathbf{y}}]_{\mathbf{x}_1,\cdots,\mathbf{x}_L} = P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$;
2. $P_{\mathbf{s}_i} \in \lambda_{\mathbf{s}}$ *for all* $i \in [L]$;
3. $P_{\mathbf{x}_i,\mathbf{s}_i,\mathbf{y}} = P_{\mathbf{x}} P_{\mathbf{s}_i|\mathbf{x}_i} W_{\mathbf{y}|\mathbf{x}_i,\mathbf{s}_i}$ *for all* $i \in [L]$.

▶ **Definition 39** (Confusability set)**.** *The $(P_{\mathbf{x}}, L)$-confusability set $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$ of a channel $\mathcal{A} = \left(\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right)$ is defined as*

$$\mathcal{K}^{\otimes L}(P_{\mathbf{x}}) \coloneqq \left\{P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{J}^{\otimes L}(P_{\mathbf{x}}) \colon P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \text{ is } L\text{-confusable}\right\}.$$

▶ Remark 40. In the above definitions, we overload the notion of confusability for types and distributions.

$$\mathcal{K}^{\otimes L}(P_{\mathbf{x}}) = \bigcup_{n=1}^{\infty} \left\{\tau_{\underline{x}_1,\cdots,\underline{x}_L} \colon (\underline{x}_1, \cdots, \underline{x}_L) \text{ is } L\text{-confusable};\ \underline{x}_i \in \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}}),\ \forall i \in [L]\right\}.$$

▶ **Definition 41** (List decodable codes). *A code $\mathcal{C} \subset \mathcal{X}^n$ is said to be $(L-1)$-*list decodable *if no size-L list is confusable, i.e., for any $\mathcal{L} \in \binom{\mathcal{C}}{L}$, $\mathcal{L}$ is non-L-confusable.*

▶ **Definition 42** (Achievable rate and list decoding capacity). *A rate R is said to be achievable under $(L-1)$-list decoding if there is an infinite sequence of $(L-1)$-list decodable codes $\{\mathcal{C}_i\}_{i\geqslant 1}$ of blocklength $n_i \in \mathbb{Z}_{>0}$ (such that $\{n_i\}$ is a non-vanishing sequence) and rate $R(\mathcal{C}) \geqslant R$.*

The $(L-1)$-list decoding capacity is defined as the maximal achievable rate.

$$C := \limsup_{n\to\infty} \max_{\substack{\mathcal{C}\subseteq\Lambda_{\underline{\mathbf{x}}} \\ (L-1)\text{-list decodable}}} R(\mathcal{C}).$$

10 List decoding capacity

▶ **Theorem 43** (List decoding capacity). *For any adversarial channel $\mathcal{A} = (\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W)$, let*[13]

$$C := \max_{P_{\mathbf{x}}\in\lambda_{\mathbf{x}}} \min_{P_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}} I(\mathbf{x};\mathbf{y}), \tag{19}$$

which can be viewed as a generalized sphere-packing bound. The mutual information is evaluated w.r.t.

$$P_{\mathbf{x},\mathbf{y}} = \left[P_{\mathbf{x}} P_{\mathbf{s}|\mathbf{x}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right]_{\mathbf{x},\mathbf{y}}.$$

Then

1. *(Achievability) For any $\delta > 0$ and sufficiently large n, there exists $\mathcal{C}$ of rate $C - \delta$ such that it can be $\mathcal{O}(1/\delta)$ list decoded.*
2. *(Converse) For any $\mathcal{C}$ of rate $C + \delta$, $\mathcal{C}$ is $2^{\Omega(n\delta)}$-list decodable.*

Proof. We follow the idea used in the proof of list decoding theorem 4 under the standard bit-flip model but conduct the calculations under our generalized setting [39].

1. (Achievability) Let $R = C - \delta$. Fix $P_{\mathbf{x}}^* \in \lambda_{\mathbf{x}}$ to be a maximizer of expression (19). Generate a random code by sampling 2^{nR} codewords independently and uniformly from $\mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}}^*)$. We will actually show that

   ▶ **Lemma 44.** *For any $\delta > 0$ and sufficiently large n, a random $P_{\mathbf{x}}^*$-constant composition code of rate $R = C - \delta$ as defined above is $\left(\frac{1+\log|\mathcal{Y}|}{\delta} - 1\right)$-list decodable with probability at least $1 - 2^{-n(1-R)}$.*

   For every $\underline{y} \in \mathcal{Y}^n$, define *conditional typical set*

   $$\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} := \left\{\underline{x} \in \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}}^*) : \exists \underline{s} \in \Lambda_{\underline{\mathbf{s}}},\ \underline{y} = W(\underline{x},\underline{s})\right\}$$

   to be the set of all $\underline{x}$ of type $P_{\mathbf{x}}^*$ that can reach $\underline{y}$ via allowable $\underline{s} \in \Lambda_{\underline{\mathbf{s}}}$. Note that $\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}$ is precisely the list of codewords around $\underline{y}$ whose size we would like to bound. In favour of proceeding calculations, we write $\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}$ in terms of types and estimate its size. We say that a type $\tau_{\mathbf{x},\mathbf{s},\mathbf{y}} \in \mathcal{P}^{(n)}(\mathcal{X}\times\mathcal{S}\times\mathcal{Y})$ is *valid* if

   a. $[\tau_{\mathbf{x},\mathbf{s},\mathbf{y}}]_{\mathbf{x}} = P_{\mathbf{x}}^*$;
   b. $[\tau_{\mathbf{x},\mathbf{s},\mathbf{y}}]_{\mathbf{s}} \in \lambda_{\mathbf{s}}$;
   c. $\tau_{\mathbf{x},\mathbf{s},\mathbf{y}} = P_{\mathbf{x}}^* \tau_{\mathbf{s}|\mathbf{x}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$.

[13] It can be easily seen that the set $\lambda_{\mathbf{s}|\mathbf{x}}$ is immediately specified given $P_{\mathbf{x}}$, $\lambda_{\mathbf{x}}$ and $\lambda_{\mathbf{s}}$.

Then it is not hard to see that

$$\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} = \bigcup_{\tau_{\mathbf{x},\mathbf{s},\mathbf{y}} \text{ valid}} \mathcal{T}_{\underline{\mathbf{x}}|\underline{y}}\left(\tau_{\mathbf{x}|\underline{y}}\right),$$

where $\tau_{\mathbf{x}|\underline{y}}$ is obtained from $\tau_{\mathbf{x},\mathbf{s},\mathbf{y}}$. Note that there is only a polynomial number of types and the volume of each $\mathcal{T}_{\underline{\mathbf{x}}|\underline{y}}\left(\tau_{\mathbf{x}|\underline{y}}\right)$ is dot equal to $2^{nH(\mathbf{x}|\mathbf{y})}$, where $H(\mathbf{x}|\mathbf{y})$ is evaluated w.r.t. $\left[\tau_{\mathbf{x},\mathbf{s},\underline{y}}\right]_{\mathbf{x},\underline{y}} = \tau_{\underline{y}}\tau_{\mathbf{x}|\underline{y}}$. Hence the volume of $\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}$ is

$$\begin{aligned}
\frac{1}{n}\log\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right| &\xrightarrow{n\to\infty} \max_{\tau_{\mathbf{x},\mathbf{s},\mathbf{y}} \text{ valid}} H(\mathbf{x}|\mathbf{y}) && (20)\\
&= \max_{\substack{P^*_{\mathbf{x}}\tau_{\mathbf{s}|\mathbf{x}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}}:\\ \left[P^*_{\mathbf{x}}\tau_{\mathbf{s}|\mathbf{x}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right]_{\mathbf{s}}\in\lambda_{\mathbf{s}}}} H(\mathbf{x}|\mathbf{y}) && (21)\\
&\to \max_{P_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}} H(\mathbf{x}|\mathbf{y}). && (22)
\end{aligned}$$

In Eqn. (20) and (21), the conditional entropy is evaluated w.r.t. $\left[\tau_{\mathbf{x},\mathbf{s},\mathbf{y}}\right]_{\mathbf{x},\mathbf{y}}$ and $\left[P^*_{\mathbf{x}}\tau_{\mathbf{s}|\mathbf{x}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right]_{\mathbf{x},\mathbf{y}}$, respectively. In Eqn. (22), the conditional entropy is evaluated w.r.t. $\left[P^*_{\mathbf{x}}P_{\mathbf{s}|\mathbf{x}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right]_{\mathbf{x},\mathbf{y}}$. This equality holds in the limit as n approaches infinity since types are asymptotically dense in distributions. Note that $\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} \subset \mathcal{T}_{\underline{\mathbf{x}}}(P^*_{\mathbf{x}})$. We have that the probability q that a random codeword $\underline{\mathbf{x}}$ is able to result in $\underline{y}$ via some admissible $\underline{s} \in \Lambda_{\mathbf{s}}$ is

$$\begin{aligned}
\frac{1}{n}\log q \coloneqq& \frac{1}{n}\log\Pr\left[\underline{\mathbf{x}} \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right]\\
=& \frac{1}{n}\log\frac{\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right|}{\left|\mathcal{T}_{\underline{\mathbf{x}}}(P^*_{\mathbf{x}})\right|} && (23)\\
\xrightarrow{n\to\infty}& \max_{P_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}} H(\mathbf{x}|\mathbf{y}) - H(\mathbf{x}) && (24)\\
=& -\max_{P_{\mathbf{x}}\in\lambda_{\mathbf{x}}}\min_{P_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}} I(\mathbf{x};\mathbf{y}) && (25)\\
=& -C.
\end{aligned}$$

Eqn. (23) follows since codewords are picked uniformly from $\mathcal{T}_{\underline{\mathbf{x}}}(P^*_{\mathbf{x}})$. Eqn. (24) is by Eqn. (22) and Eqn. (21). Eqn. (25) is by the choice of $P^*_{\mathbf{x}}$. The probability that there is a large list clustered around $\underline{y}$ is given by

$$\Pr_{\mathcal{C}}\left[\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} \cap \mathcal{C}\right| \geqslant L\right] \doteq \sum_{i=L}^{2^{nR}} \binom{2^{nR}}{i} q^i (1-q)^{2^{nR}-i}.$$

Let S_i denote the summand

$$S_i \coloneqq \binom{2^{nR}}{i} q^i (1-q)^{2^{nR}-i}.$$

Note that

$$\begin{aligned}
\frac{S_i}{S_{i+1}} =& \frac{i+1}{2^{nR}-i}\frac{1-q}{q}\\
\geqslant& \frac{2}{2^{n(C-\delta)}}\frac{1-2^{-nC}}{2^{-nC}} && (26)
\end{aligned}$$

$$
\begin{aligned}
&=2\cdot\frac{1}{2}\cdot 2^{n\delta} && (27)\\
&>1,
\end{aligned}
$$

where Eqn. (26) follows since $i \geqslant L \geqslant 1$ and Eqn. (27) follows since $1-2^{-nC} \geqslant \frac{1}{2}$ when $n \geqslant \frac{1}{C}$. The largest summand is the first term. Therefore we can bound the error probability by replacing each term with the first one.

$$
\begin{aligned}
\Pr_{\mathcal{C}}\left[\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\cap\mathcal{C}\right|\geqslant L\right] \leqslant& 2^{nR}\binom{2^{nR}}{L}q^{L}(1-q)^{2^{nR}-L}\\
\leqslant& 2^{nR}2^{nRL}2^{-nCL}\\
=& 2^{-n((L+1)\delta-C)}.
\end{aligned}
$$

Finally taking a union bound over all $\underline{y}\in\mathcal{Y}^n$, we know that the probability of list decoding error is at most

$$
\begin{aligned}
\Pr\left[\exists\underline{y}\in\mathcal{Y}^n,\ \left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\cap\mathcal{C}\right|\geqslant L\right] \leqslant& |\mathcal{Y}|^n\, 2^{-n((L+1)\delta-C)}\\
=& 2^{-n((L+1)\delta-C-\log|\mathcal{Y}|)},
\end{aligned}
$$

which is $2^{-\Omega(n)}$ if $L > \frac{1+\log|\mathcal{Y}|}{\delta}-1$. Specifically, taking $L=\frac{1+\log|\mathcal{Y}|}{\delta}$, we have that the list decoding error probability is at most $2^{-n(1+\delta-C)}=2^{-n(1-R)}$, as desired.

2. (Converse) Given any code $\mathcal{C}$ of rate $C+\delta$, choose the $\tau^*_{\mathbf{x}}\in\mathcal{P}^{(n)}(\mathcal{X})$ such that $|\mathcal{C}\cap\mathcal{T}_{\underline{\mathbf{x}}}(\tau^*_{\mathbf{x}})|$ is maximized. By Lemma 36, $R(\mathcal{C}')\asymp R(\mathcal{C})$. For this $\tau^*_{\mathbf{x}}$, choose legitimate $\tau^*_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}$ such that

$$
\tau^*_{\mathbf{s}|\mathbf{x}} := \operatorname*{argmin}_{\tau_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}} I(\mathbf{x};\mathbf{y}),
$$

where $I(\mathbf{x};\mathbf{y})$ is evaluated according to $\left[\tau^*_{\mathbf{x}}\tau_{\mathbf{s}|\mathbf{x}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right]_{\mathbf{x},\mathbf{y}}$. Now define $\tau^*_{\mathbf{x},\mathbf{s},\mathbf{y}}:=\tau^*_{\mathbf{x}}\tau^*_{\mathbf{s}|\mathbf{x}}W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$, $\tau^*_{\mathbf{x},\mathbf{y}}:=\left[\tau^*_{\mathbf{x},\mathbf{s},\mathbf{y}}\right]_{\mathbf{x},\mathbf{y}}$ and $\tau^*_{\mathbf{y}}:=\left[\tau^*_{\mathbf{x},\mathbf{y}}\right]_{\mathbf{y}}$. Over the randomness of selecting $\underline{\mathbf{y}}$ uniformly from $\mathcal{T}_{\underline{\mathbf{y}}}\left(\tau^*_{\mathbf{y}}\right)$, the average number of codewords in $\mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}$ is dot equal to

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\underline{\mathbf{y}}}\left[\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}\cap\mathcal{C}'\right|\right] &= \mathop{\mathbb{E}}_{\underline{\mathbf{y}}}\left[\sum_{\underline{x}\in\mathcal{C}'}\mathbb{1}_{\left\{\mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}\ni\underline{x}\right\}}\right]\\
&= \sum_{\underline{x}\in\mathcal{C}'}\Pr_{\underline{\mathbf{y}}}\left[\mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}\ni\underline{x}\right] && (28)\\
&= \sum_{\underline{x}\in\mathcal{C}'}\Pr_{\underline{\mathbf{y}}}\left[\mathcal{T}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}\left(\tau^*_{\mathbf{x}|\mathbf{y}}\right)\ni\underline{x}\right] && (29)\\
&= \sum_{\underline{x}\in\mathcal{C}'}\Pr_{\underline{\mathbf{y}}}\left[\tau_{\underline{x}|\underline{\mathbf{y}}}=\tau^*_{\mathbf{x}|\mathbf{y}}\right] && (30)\\
&= \sum_{\underline{x}\in\mathcal{C}'}\frac{1}{\left|\mathcal{T}_{\underline{\mathbf{y}}}\left(\tau^*_{\mathbf{y}}\right)\right|}\prod_{x\in\mathcal{X}}\binom{\tau^*_{\mathbf{x}}(x)n}{\tau^*_{\mathbf{y}}(1)n\cdot\tau^*_{\mathbf{x}|\mathbf{y}}(x|1),\cdots,\tau^*_{\mathbf{y}}(|\mathcal{Y}|)n\cdot\tau^*_{\mathbf{x}|\mathbf{y}}(x||\mathcal{Y}|)}. && (31)
\end{aligned}
$$

Eqn. (28) is linearity of expectation. Note that by our choice of $\tau^*_{\mathbf{x}}$ and $\tau^*_{\mathbf{s}|\mathbf{x}}$ (hence $\tau^*_{\mathbf{x},\mathbf{s},\mathbf{y}}$ and $\tau^*_{\mathbf{x},\mathbf{y}}$), $\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}$ only contains one type class $\mathcal{T}_{\underline{\mathbf{x}}|\underline{y}}\left(\tau^*_{\mathbf{x}|\mathbf{y}}\right)$, where $\tau^*_{\mathbf{x}|\mathbf{y}}$ is computed from $\tau^*_{\mathbf{x},\mathbf{y}}$. Eqn. (29) then follows. Eqn. (30) follows from the definition of type classes

(Definition 20). Eqn. (31) is by analyzing the sampling procedure from the first principle. The product is exactly, given $\underline{x} \in \mathcal{C}'$, the number of ways to pick $\underline{y}$ from $\mathcal{T}_{\underline{\mathbf{x}}|\underline{y}}\left(\tau^*_{\mathbf{y}}\right)$ such that $\tau_{\underline{x}|\underline{y}} = \tau^*_{\mathbf{x}|\mathbf{y}}$. We compute the exponent of the above expectation.

$$\begin{aligned}
\frac{1}{n}\log \mathop{\mathbb{E}}_{\underline{\mathbf{y}}}\left[\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} \cap \mathcal{C}'\right|\right] \xrightarrow{n\to\infty} & R' - H\left(\tau^*_{\mathbf{y}}\right) \\
& + \sum_{x\in\mathcal{X}} \tau^*_{\mathbf{x}}(x) \sum_{y\in\mathcal{Y}} \frac{\tau^*_{\mathbf{y}}(y)\tau^*_{\mathbf{x}|\mathbf{y}}(x|y)}{\tau^*_{\mathbf{x}}(x)} \log \frac{\tau^*_{\mathbf{x}}(x)}{\tau^*_{\mathbf{y}}(y)\tau^*_{\mathbf{x}|\mathbf{y}}(x|y)} && (32)\\
=& R - H\left(\tau^*_{\mathbf{y}}\right) + \sum_{x\in\mathcal{X}} \tau^*_{\mathbf{x}}(x) H(\mathbf{y}|\mathbf{x}=x) && (33)\\
=& R - H(\mathbf{y}) + H(\mathbf{y}|\mathbf{x}) && (34)\\
=& R - I(\mathbf{x};\mathbf{y}) \\
\geqslant& R - C && (35)\\
=& \delta.
\end{aligned}$$

Since codewords in the subcode $\mathcal{C}'$ are $\tau^*_{\mathbf{x}}$-constant composition, the summand in Eqn. (31) is independent of particular choices of $\underline{x}$. Eqn. (32) then follows from Stirling's approximation (Lemma 10). In Eqn. (33), $H(\mathbf{y}|\mathbf{x}=x)$ is drawn according to the conditional type

$$\tau^*_{\mathbf{y}|\mathbf{x}}(\cdot|x) = \frac{\tau^*_{\mathbf{y}}(\cdot)\tau^*_{\mathbf{x}|\mathbf{y}}(x|\cdot)}{\tau^*_{\mathbf{x}}(x)}.$$

In Eqn. (34), we pass types to distributions by the fact that types are dense in distributions asymptotically in n. $H(\mathbf{y})$ and $H(\mathbf{y}|\mathbf{x})$ are evaluated using distribution $\left[\tau^*_{\mathbf{x}} P^*_{\mathbf{s}|\mathbf{x}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right]_{\mathbf{x},\mathbf{y}}$, where

$$P^*_{\mathbf{s}|\mathbf{x}} := \mathop{\operatorname{argmin}}_{P_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}} I(\mathbf{x};\mathbf{y}),$$

and the objective function $I(\mathbf{x};\mathbf{y})$ is evaluated using $\left[\tau^*_{\mathbf{x}} P_{\mathbf{s}|\mathbf{x}} W_{\mathbf{y}|\mathbf{x}}\right]_{\mathbf{x},\mathbf{y}}$. Eqn. (35) is by the definition of C (Eqn. (19)). $\tau^*_{\mathbf{x}}$ always gives rise to mutual information no larger than the maximizer in C.

Therefore, we have shown that there exists at least one $\underline{y} \in \mathcal{Y}^n$ such that the corresponding list around $\underline{y}$ has size at least $2^{n(\delta - o(1))}$. ◀

11 List sizes of random codes

In this section, we show that, if L has order lower than $1/\delta$, then the code used in the proof of achievability (part 1) of the list decoding capacity theorem (Theorem 43) is list decodable with vanishingly small probability. This coupled with Theorem 43 implies that, for the majority (an exponentially close to 1 fraction) of random constant composition capacity-achieving (within gap δ) codes, $\Theta(1/\delta)$ is actually the *correct* order of their list sizes.

▶ **Corollary 45.** *For $\delta > 0$ and sufficiently large n, at least a $1 - 2^{-n(1-R)} - 2^{-n\delta + \frac{2}{\delta}\log\frac{1}{\delta}}$ fraction of $P^*_{\mathbf{x}}$-constant composition codes ($P^*_{\mathbf{x}}$ as defined in Eqn. (36)) of rate $R = C - \delta$ is $(L-1)$-list decodable, where $L = \Theta(1/\delta)$ lies within the following range*

$$L \in \left[\frac{C}{\delta}, \frac{1+\log|\mathcal{Y}|}{\delta}\right].$$

▶ **Theorem 46.** *For an adversarial channel* $\mathcal{A} = \left(\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, , \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right)$*, take an optimizing input distribution* $P_{\mathbf{x}}$ *which attains the list decoding capacity* C*,*

$$P_{\mathbf{x}}^{*} := \underset{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}}{\operatorname{argmax}} \min_{P_{\mathbf{s}|\mathbf{x}} \in \lambda_{\mathbf{s}|\mathbf{x}}} I(\mathbf{x}; \mathbf{y}). \tag{36}$$

For any $\delta > 0$*, for each sufficiently large blocklength* n*, sample a random code* $\mathcal{C}$ *of rate* $R = C - \delta$ *whose codewords are selected independently and uniformly from* $\mathcal{T}_{\underline{\mathbf{x}}}\left(P_{\mathbf{x}}^{*}\right)$*. Then* $\mathcal{C}$ *is* $< (C/\delta - 1)$*-list decodable with probability at most* $2^{-n\delta + \frac{2}{\delta}\log\frac{1}{\delta}}$*.*

The theorem follows from second moment calculations and generalizes similar theorems for list decodability of random error/erasure correction codes over $\mathbb{F}_q$ [26].

Proof. Let $M := 2^{nR}$. Define *typical set*

$$\mathcal{A}_{\underline{\mathbf{y}}} := \left\{W\left(\underline{x}, \underline{s}\right) \in \mathcal{Y}^n : \underline{x} \in \mathcal{T}_{\underline{\mathbf{x}}}\left(P_{\mathbf{x}}^{*}\right),\ \underline{s} \in \Lambda_{\underline{\mathbf{s}}}\right\}.$$

Put in the language of types, it can also be written as

$$\mathcal{A}_{\underline{\mathbf{y}}} = \bigcup_{\tau_{\mathbf{x},\mathbf{s},\mathbf{y}} \text{ valid}} \mathcal{T}_{\underline{\mathbf{y}}}\left(\tau_{\mathbf{y}}\right),$$

where $\tau_{\mathbf{y}} = \left[\tau_{\mathbf{x},\mathbf{s},\mathbf{y}}\right]_{\mathbf{y}}$. Define random variable W as a witness for non-list decodability of $\mathcal{C}$

$$W := \sum_{\underline{y} \in \mathcal{A}_{\underline{\mathbf{y}}}} \sum_{\{m_1, \cdots, m_L\} \in \binom{[M]}{L}} \mathbb{1}_{\left\{\left\{\underline{\mathbf{x}}_{m_1}, \cdots, \underline{\mathbf{x}}_{m_L}\right\} \subset \mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right\}}.$$

Then by Chebyshev's inequality,

$$\begin{aligned}
\Pr\left[\mathcal{C} \text{ is } (L-1)\text{-list decodable}\right] &= \Pr\left[\bigcap_{\underline{y} \in \mathcal{Y}^n} \left\{\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} \cap \mathcal{C}\right| < L\right\}\right] \qquad (37)\\
&\leqslant \Pr\left[\bigcap_{\underline{y} \in \mathcal{A}_{\underline{\mathbf{y}}}(P_{\mathbf{y}})} \left\{\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} \cap \mathcal{C}\right| < L\right\}\right]\\
&= \Pr\left[\left(\bigcup_{\underline{y} \in \mathcal{A}_{\underline{\mathbf{y}}}} \left\{\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} \cap \mathcal{C}\right| \geqslant L\right\}\right)^{c}\right]\\
&= \Pr\left[W = 0\right] \qquad (38)\\
&\leqslant \frac{\operatorname{Var}\left[W\right]}{\mathbb{E}\left[W\right]^2},
\end{aligned}$$

where Eqn. (38) follows since $W = 0$ if and only if none of the events $\left\{\left|\mathcal{A}_{\underline{\mathbf{x}}|\underline{y}} \cap \mathcal{C}\right| \geqslant L\right\}$ $(\underline{y} \in \mathcal{A}_{\underline{\mathbf{y}}})$ happens. In what follows, we will obtain an upper bound on $\operatorname{Var}\left[W\right]$ and a lower bound on $\mathbb{E}\left[W\right]$, and hence an upper bound on the probability (37).

Lower bounding $\mathbb{E}\left[W\right]$. We can get a lower bound on the expected value of W from a straightforward calculation.

$$\begin{aligned}
\mathbb{E}\left[W\right] &= \sum_{\underline{y} \in \mathcal{A}_{\underline{\mathbf{y}}}} \sum_{\{m_1, \cdots, m_L\} \in \binom{[M]}{L}} \Pr\left[\left\{\underline{\mathbf{x}}_{m_1}, \cdots, \underline{\mathbf{x}}_{m_L}\right\} \subset \mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right]\\
&= \sum_{\underline{y} \in \mathcal{A}_{\underline{\mathbf{y}}}} \sum_{\{m_1, \cdots, m_L\} \in \binom{[M]}{L}} \Pr\left[\underline{\mathbf{x}} \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right]^{L} \qquad (39)
\end{aligned}$$

$$\begin{aligned}
&\doteq \left|\mathcal{A}_{\underline{\mathbf{y}}}\right| \binom{M}{L} 2^{-nCL} && (40)\\
&\geqslant \left|\mathcal{A}_{\underline{\mathbf{y}}}\right| \left(\frac{M}{L}\right)^{L} 2^{-nCL}\\
&= \left|\mathcal{A}_{\underline{\mathbf{y}}}\right| 2^{-n\delta L - L\log L}.
\end{aligned}$$

Eqn. (39) follows since codewords are independent. Eqn. (40) is by Eqn. (25).

Upper bounding Var $[W]$. Define, for any $\underline{y} \in \mathcal{Y}^n$ and $\mathcal{L} \in \binom{[M]}{L}$,

$$\begin{aligned}
\mathbb{I}\left(\underline{y}, \mathcal{L}\right) :=& \mathbb{1}_{\left\{\{\underline{\mathbf{x}}_m\}_{m\in\mathcal{L}} \subset \mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right\}}\\
=& \prod_{m\in\mathcal{L}} \mathbb{1}_{\left\{\underline{\mathbf{x}}_m \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right\}},
\end{aligned}$$

as the indicator function of the event $\bigcap_{m\in\mathcal{L}} \left\{\underline{\mathbf{x}}_m \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{y}}\right\}$ that the list $\mathcal{L}$ is L-confusable w.r.t. $\underline{y}$.

Now the variance of W can be upper bounded as follows.

$$\begin{aligned}
\mathrm{Var}\left[W\right] =& \mathbb{E}\left[W^2\right] - \mathbb{E}\left[W\right]^2 && (41)\\
=& \sum_{\underline{y}_1,\underline{y}_2 \in \mathcal{A}_{\underline{\mathbf{y}}}} \sum_{\mathcal{L}_1,\mathcal{L}_2 \in \binom{[M]}{L}} \mathbb{E}\left[\mathbb{I}\left(\underline{y}_1, \mathcal{L}_1\right)\mathbb{I}\left(\underline{y}_2, \mathcal{L}_2\right)\right] - \mathbb{E}\left[\mathbb{I}\left(\underline{y}_1, \mathcal{L}_1\right)\right]\mathbb{E}\left[\mathbb{I}\left(\underline{y}_2, \mathcal{L}_2\right)\right] && (42)\\
\leqslant& \sum_{\substack{\mathcal{L}_1,\mathcal{L}_2 \in \binom{[M]}{L}\\ \mathcal{L}_1\cap\mathcal{L}_2 \neq \varnothing}} \sum_{\underline{y}_1,\underline{y}_2 \in \mathcal{A}_{\underline{\mathbf{y}}}} \mathbb{E}\left[\mathbb{I}\left(\underline{y}_1, \mathcal{L}_1\right)\mathbb{I}\left(\underline{y}_2, \mathcal{L}_2\right)\right] && (43)\\
=& \left|\mathcal{A}_{\underline{\mathbf{y}}}\right|^2 \sum_{\ell=1}^{L} \sum_{|\mathcal{L}_1\cap\mathcal{L}_2|=\ell} \Pr_{\underline{\mathbf{y}}_1,\underline{\mathbf{y}}_2,\mathcal{C}}\left[\mathcal{E}\right]. && (44)
\end{aligned}$$

Eqn. (41) follows from the definition of variance and Eqn. (42) follows from linearity of expectation. Note that $\mathbb{I}\left(\underline{y}_1, \mathcal{L}_1\right)$ and $\underline{I}\left(\underline{y}_2, \mathcal{L}_2\right)$ are independent if and only if $\mathcal{L}_1 \cap \mathcal{L}_2 = \varnothing$. When they are independent, the first expectation factors and the summand vanishes. The inequality (43) follows by dropping the negative term in the summand. In Eqn. (44), we rewrite the summation by randomizing the centers $\underline{y}_1, \underline{y}_2$ of the lists $\mathcal{L}_1, \mathcal{L}_2$. The probability is taken over $\underline{\mathbf{y}}_1$ and $\underline{\mathbf{y}}_2$ chosen uniformly at random from $\mathcal{A}_{\underline{\mathbf{y}}}$ and over the random code sampling procedure. We use $\mathcal{E}$ to denote the event that the lists $\mathcal{L}_1$ and $\mathcal{L}_2$ are simultaneously L-confusable w.r.t. $\underline{\mathbf{y}}_1$ and $\underline{\mathbf{y}}_2$, respectively,

$$\mathcal{E} := \bigcap_{m_1\in\mathcal{L}_1} \left\{\underline{\mathbf{x}}_{m_1} \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_1}\right\} \cap \bigcap_{m_2\in\mathcal{L}_2} \left\{\underline{\mathbf{x}}_{m_2} \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_2}\right\}.$$

It then suffices to bound $\Pr\left[\mathcal{E}\right]$. To this end, first define conditional typical set, for $\underline{x} \in \mathcal{X}^n$,

$$\begin{aligned}
\mathcal{A}_{\underline{\mathbf{y}}|\underline{x}} :=& \left\{W\left(\underline{x}, \underline{s}\right) \in \mathcal{Y}^n \colon \underline{s} \in \Lambda_{\underline{\mathbf{s}}}\right\}\\
=& \bigcup_{\tau_{\mathbf{x},\mathbf{s},\mathbf{y}} \text{ valid}} \mathcal{T}_{\underline{\mathbf{y}}}\left(\tau_{\mathbf{y}|\underline{x}}\right),
\end{aligned}$$

where $\tau_{\mathbf{y}|\underline{x}}$ is computed from $\tau_{\mathbf{x},\mathbf{s},\mathbf{y}}$ and $\tau_{\underline{x}}$, $\tau_{\mathbf{y}|\underline{x}} = \left[\tau_{\mathbf{x},\mathbf{s},\mathbf{y}}\right]_{\mathbf{x},\mathbf{y}} / \tau_{\underline{x}}$. Then define the following events in favour of bounding $\Pr\left[\mathcal{E}\right]$.

$$\mathcal{E}_1 := \left\{\underline{\mathbf{y}}_1 \in \mathcal{A}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}_m}\right\} \cap \left\{\underline{\mathbf{y}}_2 \in \mathcal{A}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}_m}\right\},$$

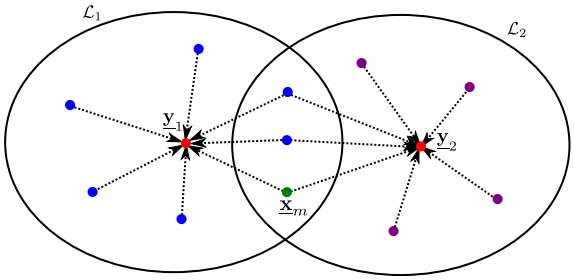


Figure 5 $\mathcal{E} \subset \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$. We upper bound $\Pr[\mathcal{E}]$ by neglecting the fact that codewords $\underline{\mathbf{x}}_i$ for $i \in (\mathcal{L}_1 \cap \mathcal{L}_2)\setminus\{m\}$ are simultaneously $\underline{\mathbf{y}}_1$-confusable and $\underline{\mathbf{y}}_2$-confusable, or equivalently, neglecting that $\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2$ should simultaneously belong to $\mathcal{A}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}_{m'}}$ for all $m' \in \mathcal{L}_1 \cap \mathcal{L}_2$, not only the particular m we have chosen.

$$\mathcal{E}_2 := \bigcap_{m_1 \in \mathcal{L}_1 \setminus \{m\}} \left\{ \underline{\mathbf{x}}_{m_1} \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_1} \right\},$$
$$\mathcal{E}_3 := \bigcap_{m_2 \in \mathcal{L}_2 \setminus \mathcal{L}_1} \left\{ \underline{\mathbf{x}}_{m_2} \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_2} \right\},$$

where $m \in \mathcal{L}_1 \cap \mathcal{L}_2$ is any message that appears in both $\mathcal{L}_1$ and $\mathcal{L}_2$. It is easy to verify that $\mathcal{E} \subset \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ (see Fig. 5). Note that $\mathcal{E}_2$ and $\mathcal{E}_3$ are independent conditioned on $\mathcal{E}_1$ since $\mathcal{L}_1\setminus\{m\}$ and $\mathcal{L}_2\setminus\mathcal{L}_1$ are disjoint. The probabilities of the above events can be computed precisely.

$$\Pr[\mathcal{E}_1] = \Pr\left[\underline{\mathbf{y}} \in \mathcal{A}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}_m}\right]^2 \tag{45}$$
$$= \left(\frac{\left|\mathcal{A}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}_m}\right|}{\left|\mathcal{A}_{\underline{\mathbf{y}}}\right|}\right)^2, \tag{46}$$

where Eqn. (45) is because $\underline{\mathbf{y}}_1$ and $\underline{\mathbf{y}}_2$ are independent, and Eqn. (46) follows since $\underline{\mathbf{y}}$ is chosen uniformly from $\mathcal{A}_{\underline{\mathbf{y}}}$. We now compute the exponent of $\Pr[\mathcal{E}]$.

$$\frac{1}{n}\log\left|\mathcal{A}_{\underline{\mathbf{y}}}\right| \xrightarrow{n\to\infty} \max_{\tau_{\mathbf{x},\mathbf{s},\mathbf{y}} \text{ valid}} H(\mathbf{y}) \tag{47}$$
$$= \max_{P_{\mathbf{s}|\mathbf{x}} \in \lambda_{\mathbf{s}|\mathbf{x}}} H(\mathbf{y}), \tag{48}$$

where in Eqn. (47) the entropy is computed w.r.t. $\tau_{\mathbf{y}} = [\tau_{\mathbf{x},\mathbf{s},\mathbf{y}}]_{\mathbf{y}}$; Eqn. (48) follows from similar calculations as done for $\mathcal{A}_{\mathbf{x}|\underline{y}}$ (Eqn. (20)) and the entropy is evaluated using $\left[P^*_{\mathbf{x}} P_{\mathbf{s}|\mathbf{x}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right]_{\mathbf{y}}$.

Similarly,

$$\frac{1}{n}\log\left|\mathcal{A}_{\underline{\mathbf{y}}|\underline{\mathbf{x}}_m}\right| \xrightarrow{n\to\infty} \max_{\tau_{\mathbf{x},\mathbf{s},\mathbf{y}} \text{ valid}} H(\mathbf{y}|\mathbf{x}) \tag{49}$$
$$= \max_{P_{\mathbf{s}|\mathbf{x}} \in \lambda_{\mathbf{s}|\mathbf{x}}} H(\mathbf{y}|\mathbf{x}), \tag{50}$$

where the conditional entropies in Eqn. (49) and (50) are evaluated w.r.t. $\tau_{\underline{x}}\tau_{\mathbf{y}|\underline{x}}$ and $\left[P^*_{\mathbf{x}} P_{\mathbf{s}|\mathbf{x}} W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right]_{\mathbf{x},\mathbf{y}}$ (since $\tau_{\underline{x}} \to P^*_{\mathbf{x}}$ as n approaches infinity), respectively. Continuing with

Eqn. (46), putting Eqn. (48) and Eqn. (50) together, we have

$$\begin{aligned}\Pr[\mathcal{E}_1] &\doteq \left(2^{n\max_{P_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}} H(\mathbf{y}|\mathbf{x})-H(\mathbf{y})}\right)^2 \\ &= 2^{-2n\min_{P_{\mathbf{s}|\mathbf{x}}\in\lambda_{\mathbf{s}|\mathbf{x}}} I(\mathbf{x};\mathbf{y})} \\ &= 2^{-2nC}, \end{aligned} \tag{51}$$

where Eqn. (51) is by the choice of $P^*_{\mathbf{x}}$ (Eqn. (36)).

We also have

$$\Pr[\mathcal{E}_2|\mathcal{E}_1] = \Pr\left[\underline{\mathbf{x}} \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_1} \middle| \mathcal{E}_1\right]^{L-1} \doteq 2^{-nC(L-1)}, \tag{52}$$

$$\Pr[\mathcal{E}_3|\mathcal{E}_1] = \Pr\left[\underline{\mathbf{x}} \in \mathcal{A}_{\underline{\mathbf{x}}|\underline{\mathbf{y}}_1} \middle| \mathcal{E}_1\right]^{L-\ell} \doteq 2^{-nC(L-\ell)}, \tag{53}$$

where Eqn. (52) and Eqn. (53) follow since $|\mathcal{L}_1| = |\mathcal{L}_2| = L$ and $|\mathcal{L}_1 \cap \mathcal{L}_2| = \ell$. We thus have, from Eqn. (51), (52) and (53), that

$$\begin{aligned}\Pr[\mathcal{E}] &\leqslant \Pr[\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3] \\ &= \Pr[\mathcal{E}_1]\Pr[\mathcal{E}_2|\mathcal{E}_1]\Pr[\mathcal{E}_3|\mathcal{E}_1] \\ &\doteq 2^{-nC(2L-\ell+1)}. \end{aligned} \tag{54}$$

Note that the number of pairs of lists $\mathcal{L}_1$ and $\mathcal{L}_2$ with intersection size ℓ is

$$\begin{aligned}\binom{M}{\ell}\binom{M-\ell}{L-\ell}\binom{M-\ell}{L-\ell} &\leqslant M^\ell M^{L-\ell} M^{L-\ell} \\ &\leqslant M^{2L-\ell}. \end{aligned} \tag{55}$$

Therefore, the variance of W can be bounded as follows.

$$\mathrm{Var}[W] \leqslant \left|\mathcal{A}_{\underline{\mathbf{y}}}\right|^2 \sum_{1\leqslant\ell\leqslant L} M^{2L-\ell} 2^{-nC(2L-\ell+1)} \tag{56}$$

$$= \left|\mathcal{A}_{\underline{\mathbf{y}}}\right|^2 2^{-nC} \sum_{1\leqslant\ell\leqslant L} 2^{-n\delta(2L-\ell)} \tag{57}$$

$$\leqslant \left|\mathcal{A}_{\underline{\mathbf{y}}}\right|^2 2^{-nC} 2^{-n\delta(2L-\ell)+\log L}, \tag{58}$$

where Eqn. (56) is by Eqn. (44), (55) and (54); Eqn. (57) is by the definition of M and the choice of R; Eqn. (58) is by replacing each term with the largest one in the summation.

Putting them together.

$$\begin{aligned}\Pr[\mathcal{C} \text{ is } (L-1)\text{-list decodable}] &\leqslant \frac{\mathrm{Var}[W]}{\mathbb{E}[W]^2} \\ &\leqslant 2^{-nC+n\delta L+(2L+1)\log L}. \end{aligned}$$

The above probability vanishes in n if $L < C/\delta$. Say $L = C/\delta - 1$, then it is at most

$$2^{-n\delta+(2(C/\delta-1)+1)\log(C/\delta-1)} \leqslant 2^{-n\delta+\frac{2}{\delta}\log\frac{1}{\delta}}.$$

◀

12 Achievability

In this section, we are going to show, via concrete random code constructions, that as long as some completely positive $(P_{\mathbf{x}}, L)$-self-coupling of order L lies outside the order-L confusability set of the channel, the $(L-1)$-list decoding capacity is positive.

Let $\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) := \mathsf{CP}^{\otimes L}_{|\mathcal{X}|} \cap \mathcal{J}^{\otimes L}(P_{\mathbf{x}})$.

▶ **Theorem 47** (Achievability). *For any given general adversarial channel $\mathcal{A} = (\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$, its $(L-1)$-list decoding capacity is positive if there is a completely positive $(P_{\mathbf{x}}, L)$-self-coupling $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathsf{CP}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}})$ outside $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$ for some $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$.*

We first state a lemma concerning the rate of a random constant composition code.

▶ **Lemma 48** (Constant composition codes). *Let $\mathcal{C} = \{\underline{\mathbf{x}}_i\}_{i=1}^{2^{nR}}$ be a random code of rate R in which each codeword is selected according to product distribution $P_{\mathbf{x}}^{\otimes n}$ independently. Let $\mathcal{C}'$ be the $P_{\mathbf{x}}$-constant composition subcode of $\mathcal{C}$, $\mathcal{C}' = \mathcal{C} \cap \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}})$. Then*

$$\Pr\left[\left|\mathcal{C}'\right| \notin (1 \pm 1/2)\frac{2^{nR}}{\nu(n)}\right] \leqslant 2\exp\left(-\frac{2^{nR}}{12\nu(n)}\right).$$

Proof. The lemma is a simple consequence of concentration of measure (Lemma 24).

$$\begin{aligned}\Pr\left[\left|\mathcal{C}'\right| \notin (1 \pm 1/2)\frac{2^{nR}}{\nu(n)}\right] &= \Pr\left[\sum_{i=1}^{2^{nR}} \mathbb{1}_{\{\tau_{\underline{\mathbf{x}}_i} = P_{\mathbf{x}}\}} \notin (1 \pm 1/2)\frac{2^{nR}}{\nu(n)}\right] \\ &\leqslant 2\exp\left(-\frac{(1/2)^2}{3}\mu\right) \qquad (59) \\ &= 2\exp\left(-\frac{2^{nR}}{12\nu(n)}\right).\end{aligned}$$

where in Eqn. (59), we note that

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{2^{nR}} \mathbb{1}_{\{\tau_{\underline{\mathbf{x}}_i} = P_{\mathbf{x}}\}}\right] &= 2^{nR}\Pr\left[\underline{\mathbf{x}} \in \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}})\right] \\ &= \frac{2^{nR}}{\nu(n)} \\ &=: \mu.\end{aligned}$$

◀

12.1 Low rate codes

Let us proceed gently. We first show that a purely random code with each entry i.i.d. w.r.t. some distribution $P_{\mathbf{x}}$ is $(L-1)$-list decodable w.h.p. as long as $P_{\mathbf{x}}^{\otimes L}$ is not L-confusable.

▶ **Lemma 49.** *For any general adversarial channel $\mathcal{A} = (\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$, if there exists a legitimate input distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ such that $P_{\mathbf{x}}^{\otimes L} \notin \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$, then the $(L-1)$-list decoding capacity of $\mathcal{A}$ is positive.*

Proof. Let $M = 2^{nR}$ for some rate R to be specified momentarily. Sample a code $\mathcal{C} = \{\underline{\mathbf{x}}_1, \cdots, \underline{\mathbf{x}}_M\}$ where each $\underline{\mathbf{x}}_i \overset{\text{i.i.d.}}{\sim} P_{\mathbf{x}}^{\otimes n}$. The expected joint type $\tau_{\underline{\mathbf{x}}_{i_1},\cdots,\underline{\mathbf{x}}_{i_L}}$ $(1 \leqslant i_1 < \cdots < i_L \leqslant M)$ of any list $\underline{\mathbf{x}}_{i_1}, \cdots, \underline{\mathbf{x}}_{i_L}$ is $P_{\mathbf{x}}^{\otimes L}$. (See Fig. 6.)

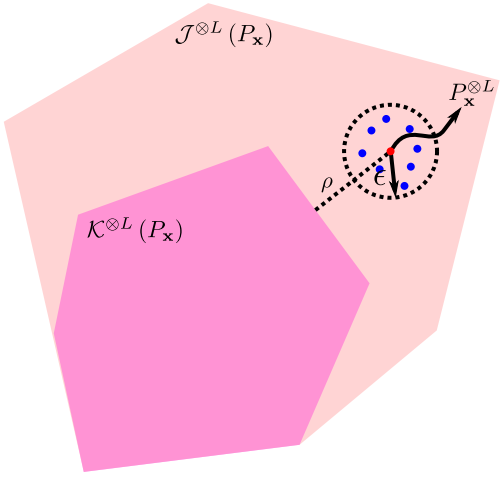


Figure 6 Low rate codes from product distribution. If the product distribution $P_{\mathbf{x}}^{\otimes L}$ is strictly separated away from $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$, then we could hope for a positive rate achieved by a random code with each entry sampled from $P_{\mathbf{x}}$. This is because w.h.p. the joint types of all (ordered) lists are contained in a $\|\cdot\|_{\text{mav}}$-ball which is completely outside the confusability set.

Let $\mathcal{C}' = \mathcal{C} \cap \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}})$ be the $P_{\mathbf{x}}$-constant composition subcode of $\mathcal{C}$. Let

$$\rho \coloneqq \inf_{P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{K}^{\otimes L}(P_{\mathbf{x}})} \left\| P_{\mathbf{x}}^{\otimes L} - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \right\|_{\text{mav}}$$

be the max-absolute-value tensor distance from the product distribution to the confusability set. Let $R = \frac{\log e}{12}\frac{\rho^2}{L} - \delta$ for some small constant $\delta > 0$. We will show that

▶ **Lemma 50.** *The random $P_{\mathbf{x}}$-constant composition code $\mathcal{C}'$ as constructed above has rate $R = \frac{\log e}{12}\frac{\rho^2}{L} - \delta$ and is $(L-1)$-list decodable with probability at least $1 - 2\exp\left(-2^{nR}/\nu(n)\right) - 2^{-n\delta + L\log|\mathcal{X}|+1}$.*

Let $\epsilon \coloneqq \rho/2$. Define error events

$$\mathcal{E}_1 \coloneqq \left\{ \left|\mathcal{C}'\right| \notin (1 \pm 1/2)\frac{2^{nR}}{\nu(n)} \right\},$$
$$\mathcal{E}_2 \coloneqq \left\{ \mathcal{C}' \text{ is not } (L-1)\text{-list decodable} \right\}.$$

By Lemma 48,

$$\Pr\left[\mathcal{E}_1\right] \leqslant 2\exp\left(-\frac{2^{nR}}{\nu(n)}\right).$$

Hence the rate R' of $\mathcal{C}'$ is asymptotically equal to R w.h.p.

By Chernoff bound,

$$\begin{aligned}
&\Pr\left[\left\|\tau_{\underline{\mathbf{x}}_{i_1},\cdots,\underline{\mathbf{x}}_{i_L}} - P_{\mathbf{x}}^{\otimes L}\right\|_{\text{mav}} \geqslant \epsilon\right] \\
=&\Pr\left[\exists (x_1,\cdots,x_L) \in \mathcal{X}^L,\ \left|\tau_{\underline{\mathbf{x}}_{i_1},\cdots,\underline{\mathbf{x}}_{i_L}}(x_1,\cdots,x_L) - P_{\mathbf{x}}(x_1)\cdots P_{\mathbf{x}}(x_L)\right| \geqslant \epsilon\right] \qquad (60)\\
\leqslant&|\mathcal{X}|^L \Pr\left[\left|\sum_{j=1}^{n} \mathbb{1}_{\left\{\left(\underline{\mathbf{x}}_{i_1}(j),\cdots,\underline{\mathbf{x}}_{i_L}(j)\right)=(x_1,\cdots,x_L)\right\}} - nP_{\mathbf{x}}(x_1)\cdots P_{\mathbf{x}}(x_L)\right| \geqslant n\epsilon\right] \qquad (61)
\end{aligned}$$

$$
\begin{aligned}
&= |\mathcal{X}|^L \Pr\left[\sum_{j=1}^{n} \mathbb{1}_{\left\{\left(\underline{\mathbf{x}}_{i_1}(j),\cdots,\underline{\mathbf{x}}_{i_L}(j)\right)=(x_1,\cdots,x_L)\right\}} \notin \left(1 \pm \frac{n\epsilon}{\mu}\right)\mu\right] && (62)\\
&\leqslant |\mathcal{X}|^L \cdot 2\exp\left(-\frac{1}{3}\left(\frac{n\epsilon}{\mu}\right)^2 \mu\right) && (63)\\
&= |\mathcal{X}|^L \cdot 2\exp\left(-\frac{n\epsilon^2}{3P_{\mathbf{x}}^{\otimes L}(x_1,\cdots,x_L)}\right) && (64)\\
&\leqslant |\mathcal{X}|^L \cdot 2\exp\left(-\frac{n}{3}\left(\frac{\rho}{2}\right)^2\right) && (65)\\
&= 2\cdot|\mathcal{X}|^L \cdot \exp\left(-\frac{\rho^2}{12}n\right).
\end{aligned}
$$

Eqn. (60) follows from the definition of max-absolute-value norm. Eqn. (61) is obtained by taking a union bound and expanding the type using definition. In Eqn. (62), we define

$$\mu \coloneqq nP_{\mathbf{x}}^{\otimes L}(x_1,\cdots,x_L),$$

which equals

$$\mathbb{E}\left[\sum_{j=1}^{n} \mathbb{1}_{\left\{\left(\underline{\mathbf{x}}_{i_1}(j),\cdots,\underline{\mathbf{x}}_{i_L}(j)\right)=(x_1,\cdots,x_L)\right\}}\right].$$

Eqn. (63) is by Chernoff bound (Lemma 24). Eqn. (64) is by the definition of μ. Eqn. (65) is by the choice of ϵ and that $P_{\mathbf{x}}^{\otimes L}(x_1,\cdots,x_L) \leqslant 1$ for any $(x_1,\cdots,x_L) \in \mathcal{X}^L$. Taking a union bound over all lists $(i_1,\cdots,i_L) \in \binom{\mathcal{M}}{L}$,

$$
\begin{aligned}
&\Pr\left[\exists\,(i_1,\cdots,i_L) \in \binom{\mathcal{M}}{L},\ \left\|\tau_{\underline{\mathbf{x}}_{i_1},\cdots,\underline{\mathbf{x}}_{i_L}} - P_{\mathbf{x}}^{\otimes L}\right\|_\infty \geqslant \epsilon\right]\\
&\leqslant \binom{M}{L} 2\cdot|\mathcal{X}|^L \cdot \exp\left(-\frac{\rho^2}{12}n\right)\\
&\leqslant 2^{-n\left(\frac{\rho^2\log e}{12} - RL\right) + L\log|\mathcal{X}|+1}.
\end{aligned}
$$

We therefore get that $\mathcal{C}$ is $(L-1)$-list decodable with probability at least $1 - 2^{-n\delta + L\log|\mathcal{X}|+1}$ as long as

$$R = \frac{\log e}{12}\frac{\rho^2}{L} - \delta.$$

Overall, we have that

$$
\begin{aligned}
\Pr[\mathcal{E}_1 \cup \mathcal{E}_2] &\leqslant \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2]\\
&\leqslant 2\exp\left(-\frac{2^{nR}}{\nu(n)}\right) + \Pr[\mathcal{C} \text{ is not } (L-1)\text{-list decodable}]\\
&\leqslant 2\exp\left(-\frac{2^{nR}}{\nu(n)}\right) + 2^{-n\delta + L\log|\mathcal{X}|+1}.
\end{aligned}
$$

◀

12.2 Random codes with expurgation

In the previous section, we only got an $(L-1)$-list decodable code of positive rate without making the effort to optimize the rate. In this section, we provide a lower bound on the $(L-1)$-list decoding capacity. It is achieved by a different code construction (random code with expurgation). However, we can only show the *existence* of such codes instead of showing that they attain the following bound w.h.p.

▶ **Lemma 51.** *The $(L-1)$-list decoding capacity of a channel $\mathcal{A}$ is at least*

$$C_{L-1} \geqslant \max_{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \min_{P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{K}^{\otimes L}(P_{\mathbf{x}})} \frac{1}{L-1} D\left(P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \| P_{\mathbf{x}}^{\otimes L}\right). \tag{66}$$

Proof. Fix any $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$ to be the maximizer of Eqn. (66). Let $M = 2^{nR}$ for some rate R to be determined. Generate a random code $\mathcal{C}$ of size $2M$ by sampling each entry of the codebook independently from $P_{\mathbf{x}}$.

For any $\underline{\mathbf{x}} \in \mathcal{C}$, by Lemma 22,

$$\Pr\left[\tau_{\underline{\mathbf{x}}} = P_{\mathbf{x}}\right] = 1/\nu(n).$$

Hence the expected number of codewords with type $P_{\mathbf{x}}$ is $2M/\nu(n)$.

For any $(\underline{\mathbf{x}}_1, \cdots, \underline{\mathbf{x}}_L) \in \binom{\mathcal{C}}{L}$,

$$\Pr\left[\tau_{\underline{\mathbf{x}}_1,\cdots,\underline{\mathbf{x}}_L} \in \mathcal{K}^{\otimes L}(P_{\mathbf{x}})\right] \doteq \sup_{P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{K}^{\otimes L}(P_{\mathbf{x}})} 2^{-nD\left(P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \| P_{\mathbf{x}}^{\otimes L}\right)},$$

by Sanov's theorem 25. Let $P^* \in \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$ be the extremizer for the above supremum. Hence the expected number of confusable lists is at most

$$\binom{2M}{L} 2^{-nD\left(P^* \| P_{\mathbf{x}}^{\otimes L}\right)} \leqslant (2M)^L 2^{-nD\left(P^* \| P_{\mathbf{x}}^{\otimes L}\right)}.$$

Pick M such that

$$(2M)^L 2^{-nD\left(P^* \| P_{\mathbf{x}}^{\otimes L}\right)} \leqslant M/\nu(n),$$

i.e.,

$$L + nRL - nD\left(P^* \| P_{\mathbf{x}}^{\otimes L}\right) \leqslant nR - \log \nu(n).$$

That is, R can be taken arbitrarily close to $\frac{1}{L-1} D\left(P^* \| P_{\mathbf{x}}^{\otimes L}\right)$.

$$\begin{aligned} R \leqslant & \frac{D\left(P^* \| P_{\mathbf{x}}^{\otimes L}\right)}{L-1} - \frac{\log \nu(n)}{(L-1)n} - \frac{L}{(L-1)n} \\ \overset{n\to\infty}{\to} & \frac{D\left(P^* \| P_{\mathbf{x}}^{\otimes L}\right)}{L-1}. \end{aligned}$$

Now, we remove all codewords of types different from $P_{\mathbf{x}}$. We also remove one codeword from each of the confusable lists. In expectation, this process reduces the size of the code by at most $2M - 2M/\nu(n)$ (due to the first expurgation) plus $(2M)^L 2^{-nD\left(P^* \| P_{\mathbf{x}}^{\otimes L}\right)} \leqslant M/\nu(n)$ (due to the second expurgation). After expurgation, we get an $(L-1)$-list decodable $P_{\mathbf{x}}$-constant composition code $\mathcal{C}'$ of size at least

$$2M - (2M/\nu(n) - 2M/\nu(n)) - M/\nu(n) = M/\nu(n).$$

The rate R' of $\mathcal{C}'$ is asymptotically the same as R.

$$\begin{aligned} R' = & R - \frac{\log \nu(n)}{n} \\ \overset{n\to\infty}{\to} & R. \end{aligned}$$

This finishes the proof. ◀

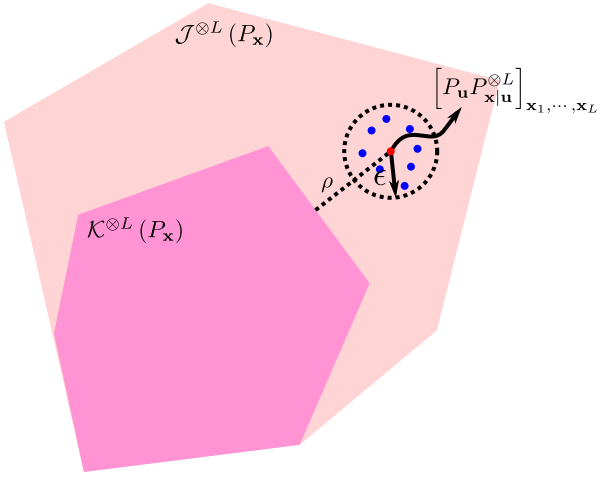


Figure 7 Low rate codes from CP distribution. If there is a CP distribution strictly outside $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$, then we can get a positive rate from random code using time-sharing. The only variation is that we divide codebook into chunks according to $P_{\mathbf{u}}$ and construct random codes of shorter length for each chunk u using distribution $P_{\mathbf{x}|\mathbf{u}=u}$.

12.3 Cloud codes

▶ **Lemma 52.** *If there is a $(P_{\mathbf{x}}, L)$-self-coupling ($P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$) $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{J}^{\otimes L}(P_{\mathbf{x}}) \backslash \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$ which can be decomposed into*

$$\begin{aligned} & P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(\underline{x}_1,\cdots,\underline{x}_L) \\ =& \sum_{u\in\mathcal{U}} P_{\mathbf{u}}(u)\, P_{\mathbf{x}|\mathbf{u}}^{\otimes L}(\underline{x}_1,\cdots,\underline{x}_L|u) \\ =& \sum_{u\in\mathcal{U}} P_{\mathbf{u}}(u) \prod_{i=1}^{L} P_{\mathbf{x}|\mathbf{u}}(\underline{x}_i|u)\,. \end{aligned}$$

for some distributions $P_{\mathbf{u}} \in \Delta(\mathcal{U})$ of finite support $|\mathcal{U}|$ and $P_{\mathbf{x}|\mathbf{u}} \in \Delta(\mathcal{X}|\mathcal{U})$ (see Fig. 7), then there exist positive rate $(L-1)$-list decodable codes.

Proof. The proof follows from a time-sharing argument combined with the previous low rate code construction (Lemma 49).

Fix R to be determined later. Sample 2^{nR} codewords in $\mathcal{C}$ independently from the following distribution. Divide each length-n codeword into $|\mathcal{U}|$ chunks $1,\cdots,|\mathcal{U}|$. For the u-th ($u \in \mathcal{U}$) chunk, sample $P_{\mathbf{u}}(u)n$ components in the chunk independently using distribution $P_{\mathbf{x}|\mathbf{u}=u}$. Let $P_{\mathbf{u},\mathbf{x}} = P_{\mathbf{u}}P_{\mathbf{x}|\mathbf{u}}$ and $P_{\mathbf{x}} = [P_{\mathbf{u},\mathbf{x}}]_{\mathbf{x}}$. Let $\mathcal{C}'$ be all codewords in $\mathcal{C}$ of type $P_{\mathbf{x}}$. (See Fig. 8.) Define

$$\rho := \inf_{P'_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{K}^{\otimes L}(P_{\mathbf{x}})} \left\| P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} - P'_{\mathbf{x},\cdots,\mathbf{x}_L} \right\|_{\text{mav}}.$$

Let

$$u^* := \operatorname*{argmin}_{u\in\mathcal{U}} P_{\mathbf{u}}(u).$$

Note that $P_{\mathbf{u}}(u^*) > 0$ since $|\mathcal{U}|$ is the *support* of $P_{\mathbf{u}}$. Let $R = \frac{P_{\mathbf{u}}(u^*)\log e}{12}\frac{\rho^2}{L} - \delta$. We will show that

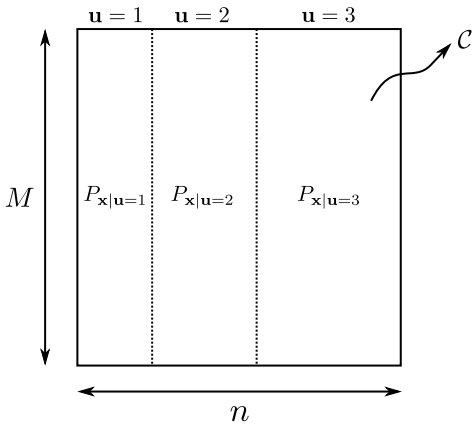


Figure 8 An example of cloud code construction in which $\mathcal{U} = \{1, 2, 3\}$. The codebook is divided into 3 chunks and symbols in the i-th chunk are sampled independently from $P_{\mathbf{x}|\mathbf{u}=i}$ $(i = 1, 2, 3)$.

▶ **Lemma 53.** *A random $P_{\mathbf{x}}$-constant composition cloud code as constructed above has rate $R = \frac{P_{\mathbf{u}}(u^*)\log e}{12}\frac{\rho^2}{L} - \delta$ and is $(L-1)$-list decodable with probability at least*

$$1 - 2\exp\left(-\frac{2^{nR}}{12\prod_{u\in\mathcal{U}}\nu(P_{\mathbf{u}}(u)n)}\right) - 2^{-n\delta + L\log|\mathcal{X}| + \log|\mathcal{U}| + 1}.$$

We write a length-n codeword as the concatenation of $|\mathcal{U}|$ chunks,

$$\underline{\mathbf{x}} = \left(\underline{\mathbf{x}}^{(1)}, \cdots, \underline{\mathbf{x}}^{(|\mathcal{U}|)}\right).$$

First we argue that w.h.p. the code $\mathcal{C}$ is almost $P_{\mathbf{x}}$-constant composition. The expected size of $\mathcal{C}'$ is

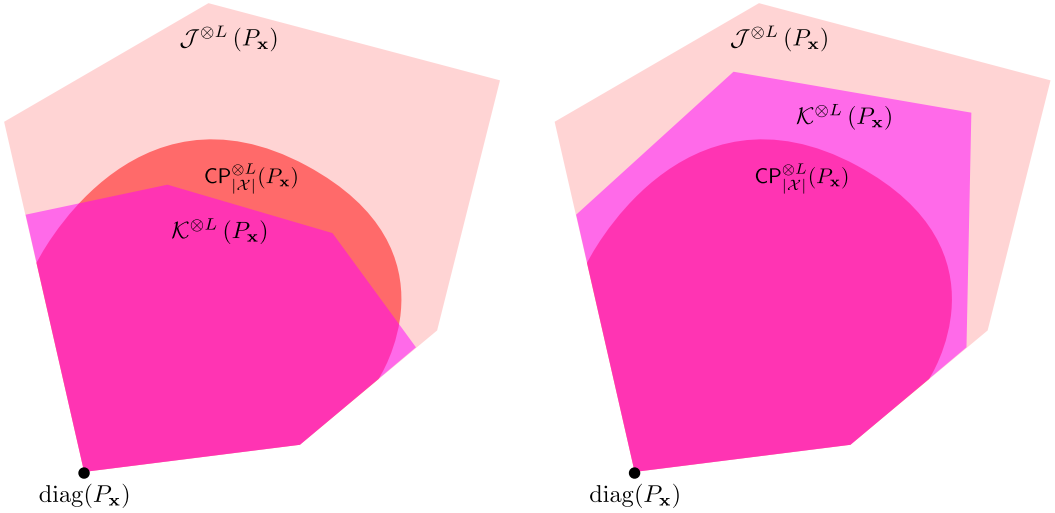
$$\begin{aligned}
\mathbb{E}\left[\left|\mathcal{C}'\right|\right] =& \mathbb{E}\left[\left|\mathcal{C} \cap \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}|\mathbf{u}})\right|\right] \\
=& \sum_{i\in[M]} \Pr\left[\underline{\mathbf{x}}_i \in \mathcal{T}_{\underline{\mathbf{x}}}(P_{\mathbf{x}|\mathbf{u}})\right] && (67)\\
=& \sum_{i\in[M]} \Pr\left[\bigcap_{u\in\mathcal{U}} \left\{\underline{\mathbf{x}}_i^{(u)} \in \mathcal{T}_{\underline{\mathbf{x}}^{(u)}}(P_{\mathbf{x}|\mathbf{u}=u})\right\}\right] \\
=& \sum_{i\in[M]} \prod_{u\in\mathcal{U}} \Pr\left[\underline{\mathbf{x}}^{(u)} \in \mathcal{T}_{\underline{\mathbf{x}}^{(u)}}(P_{\mathbf{x}|\mathbf{u}=u})\right] && (68)\\
\asymp& M\prod_{u\in\mathcal{U}} \nu(P_{\mathbf{u}}(u)n)^{-1}, && (69)
\end{aligned}$$

where Eqn. (67) is by linearity of expectation; Eqn. (68) follows since different chunks are independent; Eqn. (69) follows from Lemma 22. Then by Lemma 48

$$\Pr\left[\left|\mathcal{C}'\right| \notin (1 \pm 1/2)\mathbb{E}\left[\left|\mathcal{C}'\right|\right]\right] \leqslant 2\exp\left(-\frac{2^{nR}}{12\prod_{u\in\mathcal{U}}\nu(P_{\mathbf{u}}(u)n)}\right).$$

Secondly, for any list $1 \leqslant i_1 < \cdots < i_L \leqslant M$ of distinct ordered messages,

$$\begin{aligned}
\Pr\left[\exists u \in \mathcal{U},\ \left\|\tau_{\underline{\mathbf{x}}_{i_1}^{(u)},\cdots,\underline{\mathbf{x}}_{i_L}^{(u)}} - P_{\mathbf{x}|\mathbf{u}=u}^{\otimes L}\right\|_{\max} \geqslant \epsilon\right] &\leqslant \sum_{u\in\mathcal{U}} 2\cdot|\mathcal{X}|^L \cdot \exp\left(-\frac{\rho^2}{12}nP_{\mathbf{u}}(u)\right) && (70)\\
&\leqslant 2|\mathcal{U}||\mathcal{X}|^L \exp\left(-\frac{\rho^2}{12}nP_{\mathbf{u}}(u^*)\right), && (71)
\end{aligned}$$



(a) "Below Plotkin point", positive $(L-1)$-list decoding rate is possible. In this case, for some input distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, the slice of $P_{\mathbf{x}}$-self-coupling CP tensors is not entirely contained in the confusability set $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$.

(b) "Above Plotkin point", no positive rate for $(L-1)$-list decoding is achievable. In this case, for every input distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, the slice of $P_{\mathbf{x}}$-self-coupling CP tensors is entirely contained in the confusability set $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$.

Figure 9 A characterization of when positive rate generalized list decodable codes exist.

where the first inequality (70) follows from a union bound and same calculations as in Lemma 49. The second inequality (71) follows from the definition of u^*.

Finally, by taking another union bound over lists $\mathcal{L} \in \binom{[M]}{L}$, we get

$$\Pr\left[\exists (i_1,\cdots,i_L) \in \binom{\mathcal{M}}{L},\ \exists u \in \mathcal{U},\ \left\| \tau_{\underline{\mathbf{x}}^{(u)}_{i_1},\cdots,\underline{\mathbf{x}}^{(u)}_{i_L}} - P^{\otimes L}_{\mathbf{x}|\mathbf{u}=u} \right\|_{\mathrm{mav}} \geqslant \epsilon \right]$$
$$\leqslant 2^{-n\left(\frac{\rho^2 \log e P_{\mathbf{u}}(u^*)}{12} - RL\right) + L\log|\mathcal{X}| + \log|\mathcal{U}| + 1}.$$

Therefore, we have that the probability that the random $P_{\mathbf{x}}$-constant composition cloud code $\mathcal{C}'$ constructed above has rate $R = \frac{P_{\mathbf{u}}(u^*)\log e}{12}\frac{\rho^2}{L} - \delta$ and is $(L-1)$-list decodable with probability at least

$$1 - 2\exp\left(-\frac{2^{nR}}{12\prod_{u\in\mathcal{U}}\nu(P_{\mathbf{u}}(u)n)}\right) - 2^{-n\delta + L\log|\mathcal{X}| + \log|\mathcal{U}| + 1},$$

which completes the proof. ◀

The above lemma apparently implies Theorem 47.

13 Converse

Let $\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) := \mathsf{CP}^{\otimes L}_{|\mathcal{X}|} \cap \mathcal{J}^{\otimes L}(P_{\mathbf{x}})$ and $\mathsf{Sym}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) := \mathsf{Sym}^{\otimes L}_{|\mathcal{X}|} \cap \mathcal{J}^{\otimes L}(P_{\mathbf{x}})$.

We have shown in the previous section that if $\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) \cap \mathcal{K}^{\otimes L}(P_{\mathbf{x}})^c \neq \varnothing$, then the $(L-1)$-list decoding capacity is positive. In this section we are going to prove the converse. That is, such a condition is also necessary for positive rate being possible. Indeed, we will show that

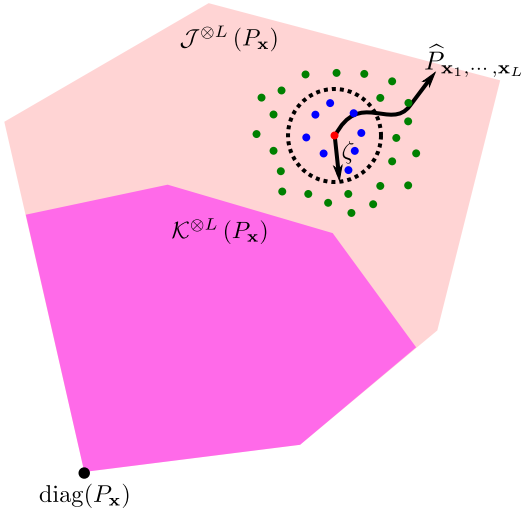


Figure 10 Equicoupled subcode extraction using hypergraph Ramsey's theorem. The union of green and blue dots represents the set of all joint types of ordered L-lists in $\mathcal{C}$. The blue dots correspond to joint types of its subcode $\mathcal{C}'$. (Note that they are all non-confusable.) They are clustered within a small ball (w.r.t. sum-absolute-value norm) centered at some distribution $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$. Since the hypergraph Ramsey number is finite, there exists such $\mathcal{C}'$ which is suitably large.

▶ **Theorem 54** (Converse). *Given a general adversarial channel* $\mathcal{A} = \left(\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x}}\right)$, *if for every admissible input distribution* $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, $\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) \subseteq \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$, *then the* $(L-1)$*-list decoding capacity of* $\mathcal{A}$ *is zero.*

13.1 Equicoupled subcode extraction

▶ **Definition 55** (Equicoupledness and ϵ-equicoupledness). *A code* $\mathcal{C}$ *is said to be* $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$-equicoupled *if for all* ordered *lists* $(\underline{x}_{i_1},\cdots,\underline{x}_{i_L}) \in \binom{\mathcal{C}}{L}$ *where* $1 \leqslant i_1 < \cdots < i_L \leqslant |\mathcal{C}|$, $\tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}} = P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$. *A code* $\mathcal{C}$ *is said to be* $(\zeta, P_{\mathbf{x}_1,\cdots,\mathbf{x}_L})$-equicoupled *if for all ordered lists* $(\underline{x}_{i_1},\cdots,\underline{x}_{i_L}) \in \binom{\mathcal{C}}{L}$, *where* $1 \leqslant i_1 < \cdots < i_L \leqslant |\mathcal{C}|$, $\left\|\tau_{\underline{x}_1,\cdots,\underline{x}_{i_L}} - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}\right\|_{\mathrm{sav}} \leqslant \epsilon$.

▶ Remark 56. The above definition can also be overloaded for sequences of random variables or their joint distributions. We say a sequence of random variables $\mathbf{w}_1,\cdots,\mathbf{w}_M$ or the joint distribution $P_{\mathbf{w}_1,\cdots,\mathbf{w}_M}$ is $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$-equicoupled (or $(\zeta, P_{\mathbf{x}_1,\cdots,\mathbf{x}_L})$-equicoupled) if every order-L marginal $P_{\mathbf{w}_{i_1},\cdots,\mathbf{w}_{i_L}}$ $(1 \leqslant i_1 < \cdots < i_L \leqslant M)$ equals (or is ζ-close to in $\|\cdot\|_{\mathrm{sav}}$) $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$.

Using the hypergraph Ramsey's theorem, we first show that any infinite sequence of codes of positive rate has an infinite sequence of subcodes which are ζ-equicoupled.

▶ **Lemma 57** (Equicoupled subcode extraction). *For any infinite sequence of codes* $\{\mathcal{C}_i\}_{i\geqslant 1}$ *of blocklengths* n_i*'s and positive rate, where* $\{n_i\}_{i\geqslant 1}$ *is an infinite increasing integer sequence, for any* $\zeta > 0$ *and any* $M \in \mathbb{Z}_{>0}$, *there is an* $N \in \mathbb{Z}_{>0}$ *such that if* $|\mathcal{C}_i| \geqslant N$ *then* $\mathcal{C}'$ *contains a subcode* $\mathcal{C}'_i$ *satisfying that*

- $|\mathcal{C}'_i| \geqslant M$;
- $\mathcal{C}'_i$ *is* $(\zeta, P_{\mathbf{x}_1,\cdots,\mathbf{x}_L})$*-equicoupled for some* $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$.

See Fig. 10.

Again, this lemma is a consequence of the hypergraph Ramsey's theorem. Let us denote by $R_c^{(m)}(n_1, \cdots, n_c)$ the smallest integer n such that the complete m-uniform hypergraph on n vertices with any c-colouring of hyperedges contains at least one of a clique of colour 1 and size n_1, ..., a clique of colour c and size n_c. It is known that $R_c^{(m)}(n_1, \cdots, n_c)$ is finite (Lemma 101), i.e., independent of the size n of the hypergraph.

Proof of Lemma 57. Recall that we assume $\mathsf{CP}_{|\mathcal{X}|}^{\otimes L}(P_{\mathbf{x}}) \cap \mathcal{K}^{\otimes L}(P_{\mathbf{x}})^c = \varnothing$. Let ρ be the gap between $\mathsf{CP}_{|\mathcal{X}|}^{\otimes L}(P_{\mathbf{x}})$ and $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$,

$$\rho := \inf_{\substack{P \in \mathsf{CP}_{|\mathcal{X}|}^{\otimes L}(P_{\mathbf{x}}) \\ P' \in \mathcal{J}^{\otimes L}(P_{\mathbf{x}}) \backslash \mathcal{K}^{\otimes L}(P_{\mathbf{x}})}} \left\| P - P' \right\|_{\mathrm{sav}}.$$

▶ **Definition 58** (ϵ-net)**.** *For a metric space $(\mathcal{X}, d)$, an ϵ-net $\mathcal{N} \subset \mathcal{X}$ is a subset which is a discrete ϵ-approximation of $\mathcal{X}$ in the sense that for any $x \in \mathcal{X}$, there is an $x' \in \mathcal{N}$ such that $d(x, x') \leqslant \epsilon$.*

We claim that

▶ **Lemma 59** (Bound on size of ϵ-net)**.** *There is an ϵ-net $\mathcal{N}$ of $\mathcal{J}^{\otimes L}(P_{\mathbf{x}}) \backslash \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$ equipped with ℓ^1 metric of size at most $\left(\frac{|\mathcal{X}|^L}{2\epsilon} + 1\right)^{|\mathcal{X}|^L}$.*

Proof. The following construction is by no means optimal, but its size has a *finite* upper bound which is enough for our purposes. Indeed, it suffices to take $\mathcal{N}$ to be the coordinate-quantization net of $\mathcal{J}^{\otimes L}(P_{\mathbf{x}}) \backslash \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$. Note that for any $P \in \mathcal{J}^{\otimes L}(P_{\mathbf{x}})$, each entry of P lies in $[0, 1]$. Take $\delta := \frac{2\epsilon}{|\mathcal{X}|^L}$. Divide $[0, 1]$ into sub-intervals of length δ (possibly except the last sub-interval that may have length less than δ). For each entry of P, there are at most $\frac{1}{\delta} + 1$ sub-intervals. Quantize each component of P to the nearest middle point of these sub-intervals. The set of all representatives whose components take values from the set of middle points of the sub-intervals form a net $\mathcal{N}$. In total, there are at most $\left(\frac{1}{\delta} + 1\right)^{|\mathcal{X}|^L}$ such representatives. For any $P \in \mathcal{J}^{\otimes L}(P_{\mathbf{x}}) \backslash \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$, let $Q_{\mathcal{N}}(P)$ denote the quantization of P using $\mathcal{N}$, i.e.,

$$Q_{\mathcal{N}}(P) := \operatorname*{argmin}_{P' \in \mathcal{N}} \left\| P - P' \right\|_{\mathrm{sav}}.$$

The quantization error is at most

$$\begin{aligned} \left\| P - Q_{\mathcal{N}}(P) \right\|_{\mathrm{sav}} &\leqslant \sum_{(x_1, \cdots, x_L) \in \mathcal{X}^L} \left| P(x_1, \cdots, x_L) - Q_{\mathcal{N}}(P)(x_1, \cdots, x_L) \right| \\ &\leqslant |\mathcal{X}|^L \frac{\delta}{2} \\ &\leqslant \epsilon. \end{aligned}$$

We thus have shown that $\mathcal{N}$ constructed as above is an ϵ-quantizer of small cardinality. ◀

Let

$$\lambda := - \sup_{\widehat{P} \in (\mathcal{J}^{\otimes L}(P_{\mathbf{x}}) \backslash \mathcal{K}^{\otimes L}(P_{\mathbf{x}})) \cap \mathsf{Sym}_{|\mathcal{X}|}^{\otimes L}(P_{\mathbf{x}})} \inf_{Q \in \mathsf{coP}_{|\mathcal{X}|}^{\otimes L}} \left\langle \widehat{P}, Q \right\rangle. \tag{72}$$

We know that CP cone and coP cone are dual (Theorem 96) in the space of symmetric tensor cone. Thus, for any non-CP *symmetric* tensor $\widehat{P} \in \mathsf{Sym}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) \backslash \mathsf{CP}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}})$, there must be a witness Q with strictly negative inner product with $\widehat{P}$. The infimum

$$\inf_{Q \in \mathsf{coP}^{\otimes L}_{|\mathcal{X}|}} \left\langle \widehat{P}, Q \right\rangle < 0.$$

λ is the absolute value of the smallest inner product among all symmetric non-CP tensors. We know that $\lambda > 0$, since $\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}})$ is *strictly* contained in $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$.

Let

$$\zeta := \frac{1}{2} \min \left\{ \rho, \frac{\lambda}{|\mathcal{X}|^L} \right\}. \tag{73}$$

Take a ζ-net of $\left(\Delta\left(\mathcal{X}^L\right), \ell^1\right)$ as constructed in Lemma 59. Such a net has cardinality at most $K := \left(\frac{|\mathcal{X}|^L}{\rho} + 1\right)^{|\mathcal{X}|^L}$.

Build an L-uniform complete hypergraph $\mathcal{H} = (\mathcal{C}, \mathcal{E})$ on $\mathcal{C}$. The vertices of $\mathcal{H}$ are codewords in $\mathcal{C}$. For every tuple $\left(\underline{x}_{i_1}, \cdots, \underline{x}_{i_L}\right) \in \binom{\mathcal{C}}{L}$ (where the indices $1 \leqslant i_1 < \cdots < i_L \leqslant |\mathcal{C}|$ are sorted in ascending order) of distinct codewords, there is a hyperedge connecting them. There are totally $\binom{|\mathcal{C}|}{L}$ hyperedges in $\mathcal{E}$. We now label hyperedges using distributions in $\mathcal{N}$. For each hyperedge $\left(\underline{x}_{i_1}, \cdots, \underline{x}_{i_L}\right) \in \mathcal{E}$, label it using the unique element $Q_{\mathcal{N}}\left(\tau_{\underline{x}_{i_1}, \cdots, \underline{x}_{i_L}}\right)$ from $\mathcal{N}$. This can be viewed as an edge colouring of $\mathcal{H}$ using at most K colours.

By hypergraph Ramsey's theorem (Theorem 101), there is a constant N such that if the size $|\mathcal{C}|$ of the hypergraph is at least N, then there is a monochromatic (each hyperedge in the sub-hypergraph has the same colour) clique $\mathcal{C}' \subset \mathcal{C}$ of size at least M. Indeed, we can take N to be the hypergraph Ramsey number $N = R^{(L)}_K(M, \cdots, M)$. By Theorem 102, there is a constant $c' > 0$ such that $N < t_L(c' \cdot K \log K)$, where $t_L(\cdot)$ is the tower function of height L. Put in another way, there exists a subcode $\mathcal{C}' \subset \mathcal{C}$ of size at least M such that for some distribution $\widehat{P}_{\mathbf{x}_1, \cdots, \mathbf{x}_L} \in \mathcal{N}$, the joint type of every ordered tuple of L distinct codewords in $\mathcal{C}'$ is ζ-close to $\widehat{P}_{\mathbf{x}_1, \cdots, \mathbf{x}_L}$. I.e., for every $\mathcal{L} = (\underline{x}_1, \cdots, \underline{x}_L) \in \binom{\mathcal{C}'}{L}$,

$$\left\| \tau_{\underline{x}_1, \cdots, \underline{x}_L} - \widehat{P}_{\mathbf{x}_1, \cdots, \mathbf{x}_L} \right\|_{\mathrm{sav}} \leqslant \zeta.$$

This completes the proof of Lemma 57. ◀

Before proceeding with the proof of converse, we first list several corollaries that directly follow from the above lemma. They are concerned with basic properties of $(\zeta, P_{\mathbf{x}_1, \cdots, \mathbf{x}_L})$-equicoupled codes.

▶ **Corollary 60.** *Any two lists of L (ordered) codewords from $\mathcal{C}'$ have joint types 2ζ close to each other in sum-absolute-value distance.*

Proof. For any $\mathcal{L}_1 = \left(\underline{x}_{i_1}, \cdots, \underline{x}_{i_L}\right)$ and $\mathcal{L}_2 = \left(\underline{x}_{j_1}, \cdots, \underline{x}_{j_L}\right)$ in $\binom{\mathcal{C}'}{L}$,

$$\begin{aligned} \left\| \tau_{\underline{x}_{i_1}, \cdots, \underline{x}_{i_L}} - \tau_{\underline{x}_{j_1}, \cdots, \underline{x}_{j_L}} \right\|_{\mathrm{sav}} &\leqslant \left\| \tau_{\underline{x}_{i_1}, \cdots, \underline{x}_{i_L}} - \widehat{P}_{\mathbf{x}_1, \cdots, \mathbf{x}_L} \right\|_{\mathrm{sav}} + \left\| \widehat{P}_{\mathbf{x}_1, \cdots, \mathbf{x}_L} - \tau_{\underline{x}_{j_1}, \cdots, \underline{x}_{j_L}} \right\|_{\mathrm{sav}} \\ &\leqslant \zeta + \zeta \\ &= 2\zeta. \end{aligned} \tag{74}$$

◀

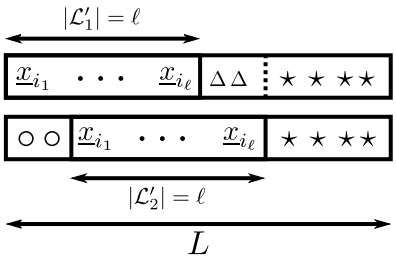


Figure 11 Two ways to complete the size-ℓ list $i_1, \cdots, i_\ell$ to size-L lists $\mathcal{L}_1, \mathcal{L}_2$, respectively. Triangles Δ, circles $\circ$ and stars $\star$ represent indices j's, k's and l's, respectively.

▶ **Corollary 61.** *Any two size-ℓ ($1 \leqslant \ell \leqslant L$) lists in $\mathcal{C}'$ have joint type 2ζ close to each other in sum-absolute-value distance, provided $|\mathcal{C}'| > 2L$.*

Proof. For any $\mathcal{L}'_1 = (\underline{x}_{i_1}, \cdots \underline{x}_{i_{L-1}})$ and $\mathcal{L}'_2 = (\underline{x}_{j_1}, \cdots, \underline{x}_{j_{L-1}})$ in $\binom{\mathcal{C}'}{L-1}$, take $\underline{x}_\iota \in \mathcal{C}' \backslash (\mathcal{L}'_1 \cup \mathcal{L}'_2)$. (This can be done as long as $|\mathcal{C}'| > 2L$.) Without loss of generality, assume $\iota > \max\{i_{L-1}, j_{L-1}\}$. Let $\mathcal{L}_1 := \mathcal{L}'_1 \cup \{\underline{x}_\iota\}, \mathcal{L}_2 := \mathcal{L}'_2 \cup \{\underline{x}_\iota\}$. We know that

$$
\begin{aligned}
2\zeta \geqslant & \left\| \tau_{\underline{x}_{i_1}, \cdots, \underline{x}_{i_{L-1}}, \underline{x}_\iota} - \tau_{\underline{x}_{j_1}, \cdots, \underline{x}_{j_{L-1}}, \underline{x}_\iota} \right\|_{\text{sav}} \\
= & \sum_{(x_1, \cdots, x_{L-1}, x) \in \mathcal{X}^L} \left| \tau_{\underline{x}_{i_1}, \cdots, \underline{x}_{i_{L-1}}, \underline{x}_\iota}(x_1, \cdots, x_{L-1}, x) - \tau_{\underline{x}_{j_1}, \cdots, \underline{x}_{j_{L-1}}, \underline{x}_\iota}(x_1, \cdots, x_{L-1}, x) \right| \\
\geqslant & \sum_{(x_1, \cdots, x_{L-1}) \in \mathcal{X}^{L-1}} \\
& \qquad \left| \sum_{x \in \mathcal{X}} \left(\tau_{\underline{x}_{i_1}, \cdots, \underline{x}_{i_{L-1}}, \underline{x}_\iota}(x_1, \cdots, x_{L-1}, x) - \tau_{\underline{x}_{j_1}, \cdots, \underline{x}_{j_{L-1}}, \underline{x}_\iota}(x_1, \cdots, x_{L-1}, x) \right) \right| \\
= & \sum_{(x_1, \cdots, x_{L-1}) \in \mathcal{X}^{L-1}} \left| \tau_{\underline{x}_{i_1}, \cdots, \underline{x}_{i_{L-1}}}(x_1, \cdots, x_{L-1}) - \tau_{\underline{x}_{j_1}, \cdots, \underline{x}_{j_{L-1}}}(x_1, \cdots, x_{L-1}) \right| \\
= & \left\| \tau_{\underline{x}_{i_1}, \cdots, \underline{x}_{i_{L-1}}} - \tau_{\underline{x}_{j_1}, \cdots, \underline{x}_{j_{L-1}}} \right\|_{\text{sav}}.
\end{aligned}
$$

Similarly we can see that Eqn. (74) holds also for size-ℓ ($\ell \leqslant L$) lists. ◀

For a subset $\mathcal{B} \subset [n]$, we let $P_{\mathbf{x}_\mathcal{B}}$ denote the marginalization of $P_{\mathbf{x}_1, \cdots, \mathbf{x}_L}$ onto the random variables indexed by elements in $\mathcal{B}$, $[P_{\mathbf{x}_1, \cdots, \mathbf{x}_L}]_{\{\mathbf{x}_i \,:\, i \in \mathcal{B}\}}$.

▶ **Corollary 62.** *For any $1 \leqslant \ell < L$ and any subsets $\mathcal{L}'_1, \mathcal{L}'_2 \in \binom{[n]}{\ell}$, $P_{\mathbf{x}_{\mathcal{L}'_1}}$ and $P_{\mathbf{x}_{\mathcal{L}'_2}}$ are 3ζ close to each other in sum-absolute-value distance, given $|\mathcal{C}'| > 2L$.*

Proof. Given two subsets $\mathcal{L}'_1, \mathcal{L}'_2 \subset [n]$ both of cardinality $\ell < L$, as long as the code size M is larger than $2L$, we can always find a tuple $1 \leqslant i_1 < \cdots < i_\ell \leqslant M$ such that it can be completed to L-tuples $\mathcal{L}_1, \mathcal{L}_2$ in two different ways

$$
\begin{aligned}
\mathcal{L}_1 = (&i_1, \cdots, i_{\ell-\ell'}, i_{\ell-\ell'+1}, \cdots, i_\ell, j_1, \cdots, j_{\ell-\ell'}, l_1, \cdots, l_{L-(2\ell-\ell')}), \\
\mathcal{L}_2 = (&k_1, \cdots, k_{\ell-\ell'}, i_1, \cdots, i'_\ell, i_{\ell'+1}, \cdots, i_\ell, l_1, \cdots, l_{L-(2\ell-\ell')}),
\end{aligned}
$$

for some $1 \leqslant k_1 < \cdots < k_{\ell-\ell'} < i_1 < \cdots < i_\ell < j_1 < \cdots < j_{\ell-\ell'} < l_1 < \cdots < l_{L-(2\ell-\ell')} \leqslant M$, where $\ell' = |\mathcal{L}'_1 \cap \mathcal{L}'_2|$. See Fig. 11. We know that

$$
\begin{aligned}
\|\tau_{\mathcal{L}_1} - P_{\mathbf{x}_1, \cdots, \mathbf{x}_L}\|_{\text{sav}} \leqslant & \zeta, \\
\|\tau_{\mathcal{L}_2} - P_{\mathbf{x}_1, \cdots, \mathbf{x}_L}\|_{\text{sav}} \leqslant & \zeta.
\end{aligned}
$$

Note that

$$\begin{aligned}
\zeta \geqslant & \left\| \tau_{\underline{x}_{\mathcal{L}_1}} - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \right\|_{\text{sav}} \\
= & \sum_{\mathcal{L}_1 \in \{0,1\}^L} \left| \tau_{\underline{x}_{\mathcal{L}_1}}(\mathcal{L}_1) - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(\mathcal{L}_1) \right| \\
\geqslant & \sum_{i_1,\cdots,i_\ell} \Bigg| \sum_{\mathcal{L}_1 \backslash \{i_1,\cdots,i_\ell\} \in \{0,1\}^{L-\ell}} \tau_{\underline{x}_{\mathcal{L}_1}}(i_1,\cdots,i_\ell,\mathcal{L}_1 \backslash \{i_1,\cdots,i_\ell\}) \\
& \qquad - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(i_1,\cdots,i_\ell,\mathcal{L}_1 \backslash \{i_1,\cdots,i_\ell\}) \Bigg| \\
\leqslant & \sum_{i_1,\cdots,i_\ell} \left| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_\ell}}(i_1,\cdots,i_\ell) - P_{\mathbf{x}_{\mathcal{L}'_1}}(i_1,\cdots,i_\ell) \right| \\
= & \left\| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_\ell}} - P_{\mathbf{x}_{\mathcal{L}'_1}} \right\|_{\text{sav}}.
\end{aligned}$$

Similarly,

$$\left\| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_\ell}} - P_{\mathbf{x}_{\mathcal{L}'_2}} \right\|_{\text{sav}} \leqslant \zeta.$$

By triangle inequality,

$$\begin{aligned}
\left\| P_{\mathbf{x}_{\mathcal{L}'_1}} - P_{\mathbf{x}_{\mathcal{L}'_2}} \right\|_{\text{sav}} \leqslant & \left\| P_{\mathbf{x}_{\mathcal{L}'_1}} - \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_\ell}} \right\|_{\text{sav}} + \left\| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_\ell}} - P_{\mathbf{x}_{\mathcal{L}'_2}} \right\|_{\text{sav}} \\
\leqslant & 2\zeta.
\end{aligned}$$

◀

▶ **Corollary 63.** *A $(\zeta, P_{\mathbf{x}_1,\cdots,\mathbf{x}_L})$-equicoupled code $\mathcal{C}'$ is $(3\zeta, P_{\mathbf{x}_1,\cdots,\mathbf{x}_\ell})$-equicoupled for any $1 \leqslant \ell \leqslant L$, as long as $|\mathcal{C}'| > 2L$.*

Proof. For any list of codewords $\underline{x}_{i_1}, \cdots, \underline{x}_{i_\ell}$, we can always find a completion of $(i_1, \cdots, i_\ell)$ to an L-tuple. Let $\mathcal{T}$ denote the set of locations of $i_1, \cdots, i_\ell$ in the completion. We know that

$$\left\| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_\ell}} - P_{\mathbf{x}_\mathcal{T}} \right\|_{\text{sav}} \leqslant \zeta.$$

By the previous corollary,

$$\begin{aligned}
\left\| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_\ell}} - P_{\mathbf{x}_1,\cdots,\mathbf{x}_\ell} \right\|_{\text{sav}} \leqslant & \left\| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_\ell}} - P_{\mathbf{x}_\mathcal{T}} \right\|_{\text{sav}} + \left\| P_{\mathbf{x}_\mathcal{T}} - P_{\mathbf{x}_1,\cdots,\mathbf{x}_\ell} \right\|_{\text{sav}} \\
\leqslant & \zeta + 2\zeta \\
= & 3\zeta.
\end{aligned}$$

◀

Now we apply the double counting trick used in the Plotkin-type bound for list decoding. We want to show that if $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is not completely positive, then any $(L-1)$-list decodable code cannot be large.

▶ **Definition 64** (Symmetry of tensors). *A tensor $T \in \mathsf{Ten}_n^{\otimes m}$ is said to be* symmetric *if its components are invariant under permutation of indices, i.e., for any $\sigma \in S_m$ and any $(t_1, \cdots, t_m) \in [n]^m$,*

$$T(t_1, \cdots, t_m) = T\left(t_{\sigma(1)}, \cdots, t_{\sigma(m)}\right).$$

The set of dimension-n order-m symmetric tensors is denoted by $\mathsf{Sym}_n^{\otimes m}$.

13.2 Symmetric case

In this subsection, assume $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is symmetric as a dimension-$|\mathcal{X}|$ order-L tensor. We are going to show that

▶ **Lemma 65** (Converse, symmetric case). *For a general adversarial channel given by $\mathcal{A} = \left(\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\right)$ and an admissible input distribution $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, if $\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) \subseteq \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$, the any $(\zeta, P_{\mathbf{x}_1,\cdots,\mathbf{x}_L})$-equicoupled $(L-1)$-list decodable code $\mathcal{C}'$ has size at most*

$$|\mathcal{C}'| \leqslant \max\left\{2(L-1), \frac{2^{L+1}L!}{\lambda}\right\},$$

where $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathsf{Sym}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) \backslash \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$ is a symmetric, non-confusable joint distribution.

Proof. Since $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathsf{Sym}^{\otimes L}_{|\mathcal{X}|}(P_{\mathbf{x}}) \backslash \mathsf{CP}^{\otimes L}_{|\mathcal{X}|}$, by duality (Theorem 96) between the CP tensor cone and coP tensor cone, there is a copositive tensor $Q \in \mathsf{coP}^{\otimes L}_{|\mathcal{X}|}$ such that $\|Q\|_{\mathrm{F}} = 1$ (by normalization) and

$$\left\langle P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}, Q \right\rangle = -\eta \tag{75}$$

for some $\eta > 0$. Note that, by definition of λ, $\eta > \lambda$. We will bound

$$\sum_{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L} \left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q \right\rangle$$

from above and below and argue that if $|\mathcal{C}'|$ is larger than some constant[14], then we get a strictly negative upper bound and a non-negative lower bound. Such a contradiction implies that no positive rate is possible for $(L-1)$-list decoding if $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is a non-CP symmetric distribution.

Upper bound.

Case when $i_1,\cdots,i_L \in [|\mathcal{C}'|]$ are not all distinct. For $i_1 \leqslant \cdots \leqslant i_L \in [|\mathcal{C}'|]$ not all distinct,

$$\left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q \right\rangle \leqslant \left\| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}} \right\|_{\mathrm{F}} \|Q\|_{\mathrm{F}} \tag{76}$$

$$\leqslant \left\| \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}} \right\|_{\mathrm{sav}} \|Q\|_{\mathrm{F}} \tag{77}$$

$$\leqslant 1. \tag{78}$$

Eqn. (76) is by Cauchy–Schwarz inequality. Eqn. (77) is because q-norm of a vector is non-increasing in q. Eqn. (78) is because a probability/type vector has one-norm 1 and Q is normalized to have F-norm 1.

Thus

$$\sum_{\substack{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L \\ \text{not all distinct}}} \left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q \right\rangle \leqslant |\mathcal{C}'|^L - \binom{|\mathcal{C}'|}{L} L!.$$

[14] Note that we will actually show that the size of the code is upper bounded by a *constant* (independent of blocklength n), not just that the rate of the code is vanishing.

Case when $i_1, \cdots, i_L \in [|\mathcal{C}'|]$ are all distinct. By Lemma 57, for any $\underline{x}_{i_1}, \cdots, \underline{x}_{i_L} \in \mathcal{C}'$ distinct,

$$\begin{aligned}\left\|\tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}} - \widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}\right\|_{\mathrm{mav}} \leqslant& \left\|\tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}} - \widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}\right\|_{\mathrm{sav}}\\ \leqslant& \zeta.\end{aligned}$$

For any $(i_1, \cdots, i_L) \in \binom{|\mathcal{C}'|}{L}$ distinct, let $\Delta_{i_1,\cdots,i_L} \coloneqq \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}} - \widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$. Immediately, $\|\Delta_{i_1,\cdots,i_L}\|_{\mathrm{mav}} \leqslant \zeta$.

Now,

$$\left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q\right\rangle = \langle \Delta_{i_1,\cdots,i_L}, Q\rangle + \left\langle \widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}, Q\right\rangle.$$

Note that

$$\begin{aligned}|\langle \Delta_{i_1,\cdots,i_L}, Q\rangle| =& \left|\sum_{(x_1,\cdots,x_L)\in\mathcal{X}^L} \Delta_{i_1,\cdots,i_L}(x_1,\cdots,x_L)Q(x_1,\cdots,x_L)\right|\\ \leqslant& \sum_{(x_1,\cdots,x_L)\in\mathcal{X}^L} |\Delta_{i_1,\cdots,i_L}(x_1,\cdots,x_L)| \qquad (79)\\ \leqslant& |\mathcal{X}|^L \cdot \zeta, \qquad (80)\end{aligned}$$

where Eqn. (79) follows from triangle inequality and $\|Q\|_{\mathrm{mav}} \leqslant \|Q\|_{\mathrm{sav}} \leqslant \|Q\|_{\mathrm{F}} = 1$.

Hence

$$\begin{aligned}\left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q\right\rangle \leqslant& -\eta + |\mathcal{X}|^L \zeta \qquad (81)\\ \leqslant& -\lambda + \frac{\lambda}{2} \qquad (82)\\ =& -\frac{\lambda}{2},\end{aligned}$$

where Eqn. (81) follows from Eqn. (75) and Eqn. (80), Eqn. (82) is by the definition of λ (Eqn. (72)) and the choice of ζ (Eqn. (73)).

Therefore,

$$\sum_{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L \text{ distinct}} \left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q\right\rangle \leqslant -\frac{\lambda}{2}\binom{|\mathcal{C}'|}{L} L!.$$

Overall,

$$\begin{aligned}\sum_{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L} \left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q\right\rangle \leqslant& |\mathcal{C}'|^L - \binom{|\mathcal{C}'|}{L} L! - \frac{\lambda}{2}\binom{|\mathcal{C}'|}{L} L!\\ <& 0 \qquad (83)\end{aligned}$$

if $|\mathcal{C}'|$ is sufficiently large. To see this, note that $p(|\mathcal{C}'|) \coloneqq |\mathcal{C}'|^L - \binom{|\mathcal{C}'|}{L} L!$ is a polynomial in $|\mathcal{C}'|$ of degree $L-1$, while $-\frac{\lambda}{2}\binom{|\mathcal{C}'|}{L} L!$ is a polynomial in $|\mathcal{C}'|$ of degree L. To give an explicit bound on $|\mathcal{C}'|$, note that the RHS of (83) equals

$$\begin{aligned}p\left(|\mathcal{C}'|\right) - \frac{\lambda}{2}|\mathcal{C}'|\left(|\mathcal{C}'|-1\right)\cdots\left(|\mathcal{C}'|-(L-1)\right) \leqslant& L\cdot(L-1)!\cdot|\mathcal{C}'|^{L-1} - \frac{\lambda}{2}\left(|\mathcal{C}'|-(L-1)\right)^L\\ =& L!\cdot|\mathcal{C}'|^{L-1} - \frac{\lambda}{2}\left(|\mathcal{C}'|-(L-1)\right)^L.\end{aligned}$$

In the above inequality, to upper bound $p\left(|\mathcal{C}'|\right)$, we replace each term of p with a monomial with the largest possible coefficient in absolute value and the largest possible degree. To make the RHS negative, we want

$$(L!)^{\frac{1}{L}}\left|\mathcal{C}'\right|^{1-\frac{1}{L}} < \left(\frac{\lambda}{2}\right)^{\frac{1}{L}}\left|\mathcal{C}'\right| - \left(\frac{\lambda}{2}\right)^{\frac{1}{L}}(L-1).$$

One can easily check that when $|\mathcal{C}'| > 2(L-1)$,

$$\frac{1}{2}\left(\frac{\lambda}{2}\right)^{\frac{1}{L}}\left|\mathcal{C}'\right| < \left(\frac{\lambda}{2}\right)^{\frac{1}{L}}\left|\mathcal{C}'\right| - \left(\frac{\lambda}{2}\right)^{\frac{1}{L}}(L-1).$$

Moreover, when $|\mathcal{C}'| > \frac{2^{L+1}L!}{\lambda}$,

$$(L!)\left|\mathcal{C}'\right|^{1-\frac{1}{L}} < \frac{1}{2}\left(\frac{\lambda}{2}\right)^{\frac{1}{L}}\left|\mathcal{C}'\right|$$

is satisfied, so is the original inequality (83).

Overall, we have that

$$\sum_{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L}\left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q\right\rangle < 0$$

as long as

$$\left|\mathcal{C}'\right| > \max\left\{2(L-1), \frac{2^{L+1}L!}{\lambda}\right\}. \tag{84}$$

Though the bound (84) is crude, it is a *constant* not depending on the blocklength n.

Lower bound.

$$\begin{aligned}
&\sum_{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L}\left\langle \tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}, Q\right\rangle\\
=&\sum_{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L}\sum_{(x_1,\cdots,x_L)\in\mathcal{X}^L}\tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}(x_1,\cdots,x_L)Q(x_1,\cdots,x_L)\\
=&\sum_{(x_1,\cdots,x_L)\in\mathcal{X}^L}\sum_{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L}\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{\left\{\underline{x}_{i_1}(j)=x_1,\cdots,\underline{x}_{i_L}(j)=x_L\right\}}Q(x_1,\cdots,x_L)\\
=&\frac{1}{n}\sum_{(x_1,\cdots,x_L)\in\mathcal{X}^L}\sum_{j=1}^{n}\sum_{(i_1,\cdots,i_L)\in[|\mathcal{C}'|]^L}\mathbb{1}_{\left\{\underline{x}_{i_1}(j)=x_1\right\}}\cdots\mathbb{1}_{\left\{\underline{x}_{i_L}(j)=x_L\right\}}Q(x_1,\cdots,x_L)\\
=&\frac{1}{n}\sum_{(x_1,\cdots,x_L)\in\mathcal{X}^L}\sum_{j=1}^{n}\left(\sum_{i\in[|\mathcal{C}'|]}\mathbb{1}_{\left\{\underline{x}_i(j)=x_1\right\}}\right)\cdots\left(\sum_{i\in[|\mathcal{C}'|]}\mathbb{1}_{\left\{\underline{x}_i(j)=x_L\right\}}\right)Q(x_1,\cdots,x_L)\\
=&\frac{|\mathcal{C}'|^L}{n}\sum_{(x_1,\cdots,x_L)\in\mathcal{X}^L}\sum_{j=1}^{n}P_{\mathbf{x}}^{(j)}(x_1)\cdots P_{\mathbf{x}}^{(j)}(x_L)Q(x_1,\cdots,x_L) \qquad (85)\\
=&\frac{|\mathcal{C}'|^L}{n}\sum_{j=1}^{n}\left\langle\left(P_{\mathbf{x}}^{(j)}\right)^{\otimes L}, Q\right\rangle\\
\geqslant&0. \qquad (86)
\end{aligned}$$

To see equality (85), let $P_{\mathbf{x}}^{(j)}$ be the empirical distribution of the j-th *column* of $\mathcal{C}'$ as a $|\mathcal{C}'| \times n$ matrix, i.e., for $x \in \mathcal{X}$,

$$P_{\mathbf{x}}^{(j)}(x) \coloneqq \frac{1}{|\mathcal{C}'|} \sum_{i=1}^{|\mathcal{C}'|} \mathbb{1}_{\{\underline{x}_i(j)=x\}}.$$

The last inequality (86) follows since $\left(P_{\mathbf{x}}^{(j)}\right)^{\otimes L}$ is a completely positive tensor.

The lower bound and the upper bound are contradicting each other, which completes the proof. ◀

13.3 Asymmetric case

In this section, we handle the asymmetric case of the converse.

▶ **Definition 66** (Asymmetry of tensors). *For a joint distribution $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \Delta\left(\mathcal{X}^L\right)$, alternatively a tensor $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathsf{Ten}_{|\mathcal{X}|}^{\otimes L}$, define its* asymmetry *as*

$$\operatorname{asymm}(P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}) \coloneqq \max_{(x_1,\cdots,x_L)\in\mathcal{X}^L} \max_{\sigma\in S_L\setminus\{\mathrm{id}\}} \left|P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_1,\cdots,x_L) - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_{\sigma(1)},\cdots,x_{\sigma(L)})\right|.$$

▶ Remark 67. If $\operatorname{asymm}(P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}) = 0$, then $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is symmetric in the sense of Definition 64.

We will show that

▶ **Lemma 68** (Converse, asymmetric case). *If $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathsf{Ten}_{|\mathcal{X}|}^{\otimes L}(P_{\mathbf{x}})$ is asymmetric as a tensor in $\mathsf{Ten}_{|\mathcal{X}|}^{\otimes L}(P_{\mathbf{x}})$ and has asymmetry α, then for any $0 < \zeta < \alpha$, any $(\zeta, P_{\mathbf{x}_1,\cdots,\mathbf{x}_L})$-equicoupled (w.r.t. max-absolute-value distance)*[15] *code $\mathcal{C}'$ has size at most*

$$|\mathcal{C}'| \leqslant \exp\left(\frac{c}{\alpha/\binom{L}{2} - \zeta}\right) + L - 2$$

for some absolute constant $c > 0$.

Lemma 68 is shown by reducing the problem, in a nontrivial way, from general values of L to $L = 2$ in which case it is known [43] that such codes cannot be large.

▶ **Lemma 69** (Reduction from general L to $L = 2$). *If $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathsf{Ten}_{|\mathcal{X}|}^{\otimes L}$ has asymmetry $\operatorname{asymm}(P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}) = \alpha$, then among the following distributions*

$$P_{\mathbf{y}_1,\mathbf{z}_1},\ P_{\mathbf{y}_2,\mathbf{z}_2},\cdots,P_{\mathbf{y}_{L-1},\mathbf{z}_{L-1}},$$

there is at least one distribution $P_{\mathbf{y}_{i^*},\mathbf{z}_{i^*}}$ ($i^* \in [L-1]$) with asymmetry at least

$$\operatorname{asymm}\left(P_{\mathbf{y}_{i^*},\mathbf{z}_{i^*}}\right) = \frac{\alpha}{\binom{L}{2}}.$$

[15] Note that ζ-equicoupledness w.r.t. sum-absolute-value distance implies ζ-equicoupledness w.r.t. max-absolute-value distance. Hence this lemma directly applies to the subcode we obtained in the previous section.

Here, for $i \in [L-1]$, $\mathbf{y}_i$ and $\mathbf{z}_i$ $(1 \leqslant i \leqslant L-1)$ are tuples of random variables defined as

$$\begin{aligned} \mathbf{y}_i &:= (\mathbf{x}_1, \cdots, \mathbf{x}_{i-1}, \mathbf{x}_i, \quad\quad \mathbf{x}_{i+2}, \cdots, \mathbf{x}_L), \\ \mathbf{z}_i &:= (\mathbf{x}_1, \cdots, \mathbf{x}_{i-1}, \quad\quad \mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \cdots, \mathbf{x}_L), \end{aligned}$$

respectively.

Proof. The proof is by contradiction. We will show that if all of $\{P_{\mathbf{y}_i,\mathbf{z}_i}\}_{1\leqslant i\leqslant L-1}$ have small asymmetry, then they do not not suffice to back propagate their asymmetry using transpositions to result in the asymmetry α of $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$. To make this intuition clear, assume, towards a contradiction, that all of the distributions $\{P_{\mathbf{y}_i,\mathbf{z}_i}\}_{1\leqslant i\leqslant L-1}$ have asymmetry strictly less than $\alpha' = \frac{\alpha}{\binom{L}{2}}$,

$$\operatorname{asymm}(P_{\mathbf{y}_i,\mathbf{z}_i}) < \frac{\alpha}{\binom{L}{2}}, \ \forall i \in [L-1]. \tag{87}$$

Assume the asymmetry of $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is witnessed by coordinates $(x_1,\cdots,x_L) \in \mathcal{X}^L$ and permutation $\pi \in S_L$, i.e.,

$$\begin{aligned} \alpha &= \left|P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_1,\cdots,x_L) - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_{\pi(1)},\cdots,x_{\pi(L)})\right| \\ &= \left|P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\pi(1)},\cdots,\mathbf{x}_{\pi(L)}}(x_1,\cdots,x_L)\right|. \end{aligned} \tag{88}$$

Note that the set of transpositions $\{\sigma_1,\cdots,\sigma_{L-1}\}$ forms a generator set of S_L, where

$$\sigma_i := \begin{pmatrix} 1 & \cdots & i-1 & i & i+1 & i+2 & \cdots & L \\ 1 & \cdots & i-1 & i+1 & i & i+2 & \cdots & L \end{pmatrix}.$$

Any permutation $\sigma \in S_L$ can be written as a product of σ_i's, $\sigma = \sigma_{i_\ell}\cdots\sigma_{i_1}$ for some positive integer ℓ and a subset of transpositions, $i_j \in [L-1]$ for each $j \in [\ell]$. Such a representation, in particular the value of ℓ, is not necessarily unique. Let

$$\ell(\sigma) := \min\{\ell \in \mathbb{Z}_{\geqslant 0} \colon \sigma = \sigma_{i_\ell}\cdots\sigma_{i_1} \text{ transposition representation}\}$$

be the *transposition length* of σ, i.e., the length of the shortest representation using product of transpositions. Let

$$\ell^* := \max_{\sigma\in S_L} \ell(\sigma).$$

We claim that $\ell^* \leqslant \binom{L}{2}$. To see this, it suffices to bound $\ell(\sigma)$ for the worst case permutation

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & L \\ L & L-1 & \cdots & 1 \end{pmatrix}.$$

The claim follows by noting that σ can be written as

$$\sigma = \prod_{j=1}^{L-1} \prod_{i=j,j-1,\cdots,1} \sigma_i, \tag{89}$$

which contains $\binom{L}{2}$ transpositions.

▶ Remark 70. A potential confusion may arise from two conflicting conventions that

1. a product is usually written from left to right, i.e.,

$$\prod_{i=1}^{\ell} \sigma_i = \sigma_1\cdots\sigma_\ell;$$

2. a composition of permutations acts like functions on an element from right to left, i.e., for $\sigma, \pi \in S_L$ and $i \in [L]$,

$$(\sigma\pi)(i) = \sigma(\pi(i)).$$

With this kept in mind, the representation in Eqn. (89) should be understood as

$$\sigma = (\sigma_1)(\sigma_2\sigma_1)\cdots(\sigma_{L-2}\cdots\sigma_2\sigma_1)(\sigma_{L-1}\cdots\sigma_2\sigma_1).$$

The product in the $(L-1)$-st parenthesis (from left to right) moves L in the initial sequence $(L, L-1, \cdots, 1)$ to the L-th position; the product in the $(L-2)$-nd parenthesis moves $L-1$ to the $(L-1)$-st position; ...; the permutation σ_1 in the 1-st parenthesis moves 2 to the 2-st position, and automatically 1 is in the 1-st position. We get the target sequence $(1, 2, \cdots, L)$. We can write

$$\pi = \prod_{j=\ell,\ell-1,\cdots,1} \sigma_{i_j}, \tag{90}$$

for some $\ell \leqslant \ell^* \leqslant \binom{L}{2}$.

Our assumption Eqn. (87) implies that, for any $(x_1, \cdots, x_L) \in \mathcal{X}^L$ and any transposition σ_i,

$$\begin{aligned}
&\left|P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\sigma_i(1)},\cdots,\mathbf{x}_{\sigma_i(L)}}(x_1,\cdots,x_L)\right| \\
=&\left|P_{\mathbf{x}_1,\cdots,\mathbf{x}_{i-1},\mathbf{x}_i,\mathbf{x}_{i+1},\mathbf{x}_{i+2},\cdots,\mathbf{x}_L}(x_1,\cdots,x_L) - P_{\mathbf{x}_1,\cdots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\mathbf{x}_i,\mathbf{x}_{i+2},\cdots,\mathbf{x}_L}(x_1,\cdots,x_L)\right| \\
=&\left|P_{(\mathbf{x}_1,\cdots,\mathbf{x}_{i-1},\mathbf{x}_i,\mathbf{x}_{i+2},\cdots,\mathbf{x}_L),(\mathbf{x}_1,\cdots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\mathbf{x}_{i+2},\cdots,\mathbf{x}_L)}\begin{pmatrix}(x_1,\cdots,x_{i-1},x_i,x_{i+2},\cdots,x_L),\\(x_1,\cdots,x_{i-1},x_{i+1},x_{i+2},\cdots,x_L)\end{pmatrix}\right. \\
&\quad\left. - P_{(\mathbf{x}_1,\cdots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\mathbf{x}_{i+2},\cdots,\mathbf{x}_L),(\mathbf{x}_1,\cdots,\mathbf{x}_{i-1},\mathbf{x}_i,\mathbf{x}_{i+2},\cdots,\mathbf{x}_L)}\begin{pmatrix}(x_1,\cdots,x_{i-1},x_i,x_{i+2},\cdots,x_L),\\(x_1,\cdots,x_{i-1},x_{i+1},x_{i+2},\cdots,x_L)\end{pmatrix}\right| \\
=&\left|P_{\mathbf{y}_i,\mathbf{z}_i}(y,z) - P_{\mathbf{z}_i,\mathbf{y}_i}(y,z)\right| \\
=&\left|P_{\mathbf{y}_i,\mathbf{z}_i}(y,z) - P_{\mathbf{y}_i,\mathbf{z}_i}(z,y)\right| \\
<&\alpha',
\end{aligned} \tag{91}$$

where

$$\begin{aligned}
y &:= (x_1, \cdots, x_{i-1}, x_i, x_{i+2}, \cdots, x_L), \\
z &:= (x_1, \cdots, x_{i-1}, x_{i+1}, x_{i+2}, \cdots, x_L).
\end{aligned}$$

Now

$$\begin{aligned}
\alpha =& \left|P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\pi(1)},\cdots,\mathbf{x}_{\pi(L)}}(x_1,\cdots,x_L)\right| && (92)\\
\leqslant& \left|P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_1}(L)}}(x_1,\cdots,x_L)\right| \\
&+ \left|P_{\mathbf{x}_{\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_1}(L)}}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\pi(1)},\cdots,\mathbf{x}_{\pi(L)}}(x_1,\cdots,x_L)\right| && (93)\\
<&\alpha' + \left|P_{\mathbf{x}_{\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_1}(L)}}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\pi(1)},\cdots,\mathbf{x}_{\pi(L)}}(x_1,\cdots,x_L)\right| && (94)\\
\leqslant&\alpha' + \left|P_{\mathbf{x}_{\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_1}(L)}}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\sigma_{i_2}\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_2}\sigma_{i_1}(L)}}(x_1,\cdots,x_L)\right| \\
&+ \left|P_{\mathbf{x}_{\sigma_{i_2}\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_2}\sigma_{i_1}(L)}}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\pi(1)},\cdots,\mathbf{x}_{\pi(L)}}(x_1,\cdots,x_L)\right| && (95)\\
<&2\alpha' + \left|P_{\mathbf{x}_{\sigma_{i_2}\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_2}\sigma_{i_1}(L)}}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\pi(1)},\cdots,\mathbf{x}_{\pi(L)}}(x_1,\cdots,x_L)\right| && (96)
\end{aligned}$$

$$\cdots$$

$$\leqslant(\ell-1)\alpha' + \left|P_{\mathbf{x}_{\sigma_{i_{\ell-1}}\cdots\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_{\ell-1}}\cdots\sigma_{i_1}(L)}}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\pi(1)},\cdots,\mathbf{x}_{\pi(L)}}(x_1,\cdots,x_L)\right| \tag{97}$$

$$=(\ell-1)\alpha' + \Big|P_{\mathbf{x}_{\sigma_{i_{\ell-1}}\cdots\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_{\ell-1}}\cdots\sigma_{i_1}(L)}}(x_1,\cdots,x_L) - P_{\mathbf{x}_{\sigma_{i_\ell}\sigma_{i_{\ell-1}}\cdots\sigma_{i_1}(1)},\cdots,\mathbf{x}_{\sigma_{i_\ell}\sigma_{i_{\ell-1}}\cdots\sigma_{i_1}(L)}}(x_1,\cdots,x_L)\Big| \tag{98}$$

$$<\ell\alpha' \tag{99}$$

$$\leqslant\binom{L}{2}\alpha'$$

$$=\alpha. \tag{100}$$

1. Eqn. (92) follows from Eqn. (88).
2. Eqn. (93), (95), etc., are by triangle inequality.
3. Eqn. (94), (96), (99), etc. , are by Eqn. (91).
4. Eqn. (97) is by recursively applying the previous calculations.
5. Eqn. (98) is by the transposition representation of π (Eqn. (90)).
6. Eqn. (100) is by the choice of α'.

We reach a contradiction that α is strictly less than itself. This finishes the proof. ◀

Next, we show the key lemma 68 in this section. Note that, according to the statement, Lemma 68 is independent of the channel that the code $\mathcal{C}'$ is used for. Hence we will directly prove the random variable version of this lemma which is concerned with fundamental properties of joint distributions. If the joint distribution of a sequence of random variables has all of its size-L marginals being ζ-close to some *asymmetric* distribution, then such a sequence cannot be infinitely long. We will prove a finite upper bound on the length of the sequence by reducing this problem from the general $L > 2$ case to the $L = 2$ case. In the $L = 2$ case, prior work [43] shows that this is indeed the case.

▶ **Lemma 71** (Converse, asymmetric case, $L = 2$ [43]). *Assume $P_{\mathbf{x}_1,\mathbf{x}_2} \in \Delta(\mathcal{X}^2)$ has asymmetry* $\operatorname{asymm}(P_{\mathbf{x}_1,\mathbf{x}_2}) = \alpha$. *Let $\mathbf{w}_1,\cdots,\mathbf{w}_M$ be a sequence of M random variables supported on $\mathcal{X}$ such that for every $1 \leqslant j_1 < j_2 \leqslant M$,*

$$\left\|P_{\mathbf{w}_{j_1},\mathbf{w}_{j_2}} - P_{\mathbf{x}_1,\mathbf{x}_2}\right\|_{\text{mav}} \leqslant \zeta.$$

for some $0 < \zeta < \alpha$. Then

$$M \leqslant \exp\left(\frac{c}{\alpha-\zeta}\right)$$

for some universal constant $c > 0$.

We are now ready to prove the restated version of Lemma 68.

▶ **Lemma 72** (Converse, asymmetric case, general L). *If a joint distribution $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \Delta\left(\mathcal{X}^L\right)$ has asymmetry* $\operatorname{asymm}(P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}) = \alpha$, *and a sequence of M random variables $\mathbf{w}_1,\cdots,\mathbf{w}_M$ supported on $\mathcal{X}$ satisfies that for any $1 \leqslant j_1 < \cdots < j_L \leqslant M$,*

$$\left\|P_{\mathbf{w}_{j_1},\cdots,\mathbf{w}_{j_L}} - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}\right\|_{\text{mav}} \leqslant \zeta. \tag{101}$$

Then

$$M \leqslant \exp\left(\frac{c}{\alpha/\binom{L}{2} - \zeta}\right) + L - 2$$

for some universal constant $c > 0$*.*

Proof. Construct the following $L-1$ sequences $\left\{\mathbf{v}^{(i)}\right\}_{1\leqslant i\leqslant L-1}$ of random variables, each of which has length $M-L+2$,

$$\begin{array}{lclllll} \mathbf{v}^{(1)} & = & (\mathbf{v}_1^{(1)}, & \mathbf{v}_2^{(1)}, & \cdots, & \mathbf{v}_{M-L+2}^{(1)}), \\ \mathbf{v}^{(2)} & = & (\mathbf{v}_2^{(2)}, & \mathbf{v}_2^{(2)}, & \cdots, & \mathbf{v}_{M-L+3}^{(2)}), \\ \cdots & & & & & \\ \mathbf{v}^{(L-1)} & = & (\mathbf{v}_{L-1}^{(L-1)}, & \mathbf{v}_2^{(1)}, & \cdots, & \mathbf{v}_M^{(L-1)}). \end{array}$$

For $1 \leqslant i \leqslant L-1$ and $i \leqslant j \leqslant M-L+i+1$, $\mathbf{v}_j^{(i)}$ is defined as atuple

$$\mathbf{v}_j^{(i)} := \left(\mathbf{w}_1, \cdots, \mathbf{w}_{i-1}, \mathbf{w}_j, \mathbf{w}_{M-L+i+2}, \cdots, \mathbf{w}_M\right).$$

Then, for any

$$\begin{array}{lclllllll} v_1 & := & (x_1, & \cdots, & x_{i-1}, & x_i, & & x_{i+2}, & \cdots, & x_L) \in \mathcal{X}^{L-1}, \\ v_2 & := & (x_1, & \cdots, & x_{i-1}, & & x_{i+1}, & x_{i+2}, & \cdots, & x_L) \in \mathcal{X}^{L-1}, \end{array}$$

and $i \leqslant j_1 < j_2 \leqslant M-L+i+1$, we have

$$\begin{aligned} &\left|P_{\mathbf{v}_{j_1}^{(i)},\mathbf{v}_{j_2}^{(i)}}(v_1, v_2) - P_{\mathbf{y}_i,\mathbf{z}_i}(v_1, v_2)\right| \\ =&\Bigg|P_{(\mathbf{w}_1,\cdots,\mathbf{w}_{i-1},\mathbf{w}_{j_1},\mathbf{w}_{M-L+i+2},\cdots,\mathbf{w}_M),(\mathbf{w}_1,\cdots,\mathbf{w}_{i-1},\mathbf{w}_{j_2},\mathbf{w}_{M-L+i+2},\cdots,\mathbf{w}_M)} \\ &\qquad\qquad\left(\begin{array}{c}(x_1,\cdots,x_{i-1},x_i,x_{i+2},\cdots,x_L),\\(x_1,\cdots,x_{i-1},x_{i+1},x_{i+2},\cdots,x_L)\end{array}\right) \\ &-P_{(\mathbf{x}_1,\cdots,\mathbf{x}_{i-1},\mathbf{x}_i,\mathbf{x}_{i+2},\cdots,\mathbf{x}_L),(\mathbf{x}_1,\cdots,\mathbf{x}_{i-1},\mathbf{x}_{i+1},\mathbf{x}_{i+2},\cdots,\mathbf{x}_L)}\left(\begin{array}{c}(x_1,\cdots,x_{i-1},x_i,x_{i+2},\cdots,x_L),\\(x_1,\cdots,x_{i-1},x_{i+1},x_{i+2},\cdots,x_L)\end{array}\right)\Bigg| \\ =&\Big|P_{\mathbf{w}_1,\cdots,\mathbf{w}_{i-1},\mathbf{w}_{j_1},\mathbf{w}_{j_2},\mathbf{w}_{M-L+i+2},\cdots,\mathbf{w}_M}(x_1,\cdots,x_{i-1},x_i,x_{i+1},x_{i+2},\cdots,x_L) \\ &\qquad\qquad - P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_1,\cdots,x_L)\Big| \\ \leqslant&\zeta, \end{aligned}$$

by the assumption Eqn. (101). Therefore, all sequences $\mathbf{v}^{(i)}$'s are $(\zeta, P_{\mathbf{y}_i,\mathbf{z}_i})$-equicoupled, $1 \leqslant i \leqslant L-1$.

Since $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is α-asymmetric, by Lemma 69, at least one of the distributions $P_{\mathbf{y}_i,\mathbf{z}_i}$'s ($1 \leqslant i \leqslant L-1$) is at least α'-asymmetric ($\alpha' = \alpha/\binom{L}{2}$). Without loss of generality, assume $P_{\mathbf{y}_{i_0},\mathbf{z}_{i_0}}$ is $\geqslant \alpha'$-asymmetric. Then the i_0-th sequence $\mathbf{v}^{(i_0)}$ is short by Lemma 71,

$$M - L + 2 \leqslant \exp\left(\frac{c}{\alpha' - \zeta}\right),$$

for some universal constant $c > 0$. Hence

$$M \leqslant \exp\left(\frac{c}{\alpha/\binom{L}{2} - \zeta}\right) + L - 2,$$

which finishes the proof. ◀

▶ Remark 73 (Asymmetric but projectively symmetric tensors). Lemma 69 does not follow from naïvely marginalizing an asymmetric distribution $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ and hoping that $P_{\mathbf{x}_i,\mathbf{x}_j}$ is asymmetric for some $1 \leqslant i < j \leqslant L$. Just like there exist asymmetric matrices (self-couplings) with the same column sum and row sum, we should not expect that the asymmetry of a tensor is preserved under projections.

We say that a tensor $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathsf{Ten}_{|\mathcal{X}|}^{\otimes L}$ is *ℓ-projectively symmetric* ($1 \leqslant \ell < L$) if all of its order-ℓ projections are symmetric, i.e., for any $1 \leqslant i_1 < \cdots < i_\ell \leqslant L$,

$$P_{\mathbf{x}_{i_1},\cdots,\mathbf{x}_{i_\ell}} \coloneqq [P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}]_{\mathbf{x}_{i_1},\cdots,\mathbf{x}_{i_\ell}} \in \mathsf{Ten}_{|\mathcal{X}|}^{\otimes \ell}$$

is symmetric.

One can easily verify the following facts.

▶ **Lemma 74.** *Let $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ be a tensor of dimension $|\mathcal{X}|$ and order L.*

1. *If $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is ℓ-projectively symmetric ($1 \leqslant \ell < L$), then all of its order-ℓ' ($1 \leqslant \ell' < \ell$) marginals are the same.*
2. *If $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is ℓ-projectively symmetric ($1 \leqslant \ell < L$), then it is also ℓ'-projectively symmetric for any $1 \leqslant \ell' < \ell$.*
3. *A symmetric tensor $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is also ℓ-projectively symmetric for all $1 \leqslant \ell < L$. In particular, it is a self-coupling, i.e., $P_{\mathbf{x}_i}$ is the same for all $i \in [L]$.*

We provide an example showing that the asymmetry of a tensor cannot be recovered from all of its lower order projections. That is, there is an asymmetric tensor with every projection of one less order being symmetric.

We now construct a concrete example. In order for a dimension-2 order-3 tensor $T\colon [2]^3 \to \mathbb{R}$ to be symmetric, it has to satisfy the following system $\mathcal{E}_1$ of linear equations,

$$t_{112} = t_{121}, \quad t_{121} = t_{211}, \quad t_{212} = t_{122}, \quad t_{122} = t_{221}.$$

where $t_{ijk} \coloneqq T(i,j,k)$ for $i,j,k \in [2]$. On the other hand, for it to be projectively symmetric, it has to satisfy the following system $\mathcal{E}_2$ of linear equations,

$$\begin{aligned} t_{122} + t_{121} &= t_{212} + t_{211}, \\ t_{112} + t_{122} &= t_{211} + t_{221}, \\ t_{121} + t_{221} &= t_{112} + t_{212}. \end{aligned}$$

Additionally, for T to represent a joint distribution, all entries should be non-negative and sum up to one. Note that $\mathcal{E}_2$ is a *less determined* system than $\mathcal{E}_1$, which means that we should be able to find a solution to $\mathcal{E}_2$ which does not satisfy $\mathcal{E}_1$.

Indeed, consider the following explicit example of $T \in \mathsf{Ten}_2^{\otimes 3}$. (See Fig. 12.)

$$\begin{aligned} t_{111} &= \frac{1}{60}, \quad t_{121} = \frac{1}{4}, \quad t_{112} = \frac{1}{6}, \quad t_{122} = \frac{1}{20}, \\ t_{211} &= \frac{1}{60}, \quad t_{221} = \frac{1}{5}, \quad t_{212} = \frac{17}{60}, \quad t_{222} = \frac{1}{60}. \end{aligned}$$

It is asymmetric but projectively symmetric. Note that T is forced to have multiple witnesses of asymmetry due to its projective symmetry. Indeed,

$$\begin{aligned} t_{121} - t_{112} &= t_{212} - t_{221} = \frac{5}{60}, \\ t_{121} - t_{211} &= t_{212} - t_{122} = \frac{14}{60}, \\ t_{112} - t_{211} &= t_{221} - t_{122} = \frac{9}{60}. \end{aligned}$$

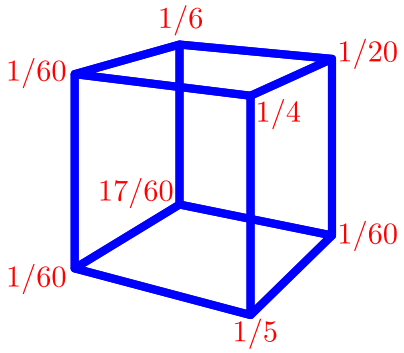


Figure 12 An asymmetric tensor $T \in \mathsf{Ten}_2^{\otimes 3}$ that is 2-projectively symmetric.

Therefore $\operatorname{asymm}(T) = \frac{14}{60} = \frac{7}{30}$, given by $t_{121} - t_{211}$ and $t_{212} - t_{122}$. All of its order-2 projections are given by

$$\begin{bmatrix} \frac{11}{60} & \frac{3}{10} \\ \frac{3}{10} & \frac{13}{60} \end{bmatrix}, \quad \begin{bmatrix} \frac{4}{15} & \frac{13}{60} \\ \frac{13}{60} & \frac{3}{10} \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{30} & \frac{9}{20} \\ \frac{9}{20} & \frac{1}{15} \end{bmatrix}.$$

All of their margins are equal to $\begin{bmatrix} \frac{29}{60} \\ \frac{31}{60} \end{bmatrix}$.

In general, for any dimension-d order-L tensor, such examples can always be constructed due to the gap of degrees of freedom between the homogeneous linear systems $\mathcal{E}_1$ and $\mathcal{E}_2$.

14 Rethinking the converse

14.1 A cheap converse

If for a general $\mathcal{A} = (\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}})$, for every $P_{\mathbf{x}} \in \lambda_{\mathbf{x}}$, the confusability set is a halfspace defined by a single linear constraint

$$\mathcal{K}^{\otimes L}(P_{\mathbf{x}}) \coloneqq \left\{ P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{J}^{\otimes L}(P_{\mathbf{x}}) \colon \langle P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}, C \rangle \leqslant b \right\},$$

for some tensor $C \in \mathsf{Ten}_{|\mathcal{X}|}^{\otimes L}$ and constant b, then the converse can be significantly simplified. In particular, we do not have to handle symmetric and asymmetric cases separately. We describe the proof idea below.

Proof. The proof essentially follow from the following observation. For any asymmetric $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$, given any $P_{\mathbf{x}}$-constant composition $(\zeta, P_{\mathbf{x}_1,\cdots,\mathbf{x}_L})$-equicoupled code $\mathcal{C} = \{\underline{x}_i\}_{i=1}^M$ in $\mathcal{X}^n$ of size M, we can construct a code $\mathcal{C}' = \{\underline{x}'_i\}_{i=1}^M$ in $\mathcal{X}^{n \cdot M!}$ of the same size which is *symmetric*. Indeed, we can permute the rows of $\mathcal{C}$ using $\sigma \in S_M$ and juxtapose all possible ($M!$ of them in total) such row-permuted codes $\sigma(\mathcal{C})$. (See Fig. 13.) The resulting code $\mathcal{C}'$ is actually not only L-wise approximately equicoupled, but M-wise exactly equicoupled! For *any* $L \in [M]$ and any L-sized (not necessarily ordered) subset $\{i_1, \cdots, i_L\}$ of $[M]$, the joint type of $\underline{x}'_{i_1}, \cdots, \underline{x}'_{i_L}$ is *exactly* equal to

$$\tau_{\underline{x}'_{i_1},\cdots,\underline{x}'_{i_L}} = \frac{1}{\binom{M}{L}} \sum_{\{i_1,\cdots,i_L\} \in \binom{[M]}{L}} \frac{1}{L!} \sum_{\sigma \in S_L} \tau_{\underline{x}_{\sigma(i_1)},\cdots,\underline{x}_{\sigma(i_L)}},$$

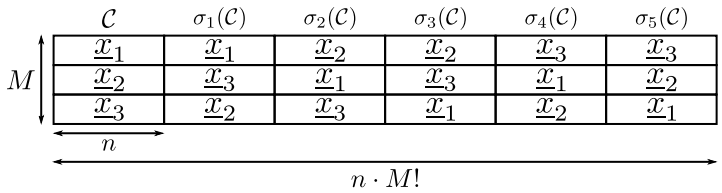


Figure 13 Construction of $\mathcal{C}'$ by permuting rows of $\mathcal{C} = \{\underline{x}_1, \underline{x}_2, \underline{x}_3\}$ using $\sigma \in S_3$ (where $S_3 = \{\mathrm{id}, \sigma_1, \cdots, \sigma_5\}$) and juxtaposing all $\sigma(\mathcal{C})$ (6 of them in total) together.

which is *symmetric* and independent of the choice of the list $(i_1, \cdots, i_L)$ (hence let us denote it by $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$). In particular, letting $L = M$, we get that

$$\tau_{\underline{x}'_1,\cdots,\underline{x}'_M} = \frac{1}{M!} \sum_{\sigma \in S_M} \tau_{\underline{x}_{\sigma(1)},\cdots,\underline{x}_{\sigma(M)}}.$$

To see the above claims, note that if we juxtapose two pairs of codewords $(\underline{x}_1, \underline{x}_2)$ and $(\underline{x}'_1, \underline{x}'_2)$, we get a pair of longer codewords $(\widetilde{\underline{x}}_1, \widetilde{\underline{x}}_2) \coloneqq (\underline{x}_1 \circ \underline{x}'_1, \underline{x}_2 \circ \underline{x}'_2)$ (where $\circ$ denotes concatenation) with joint type

$$\tau_{\widetilde{\underline{x}}_1,\widetilde{\underline{x}}_2} = \frac{1}{2}(\tau_{\underline{x}_1,\underline{x}_2} + \tau_{\underline{x}'_1,\underline{x}'_2}).$$

This still holds if two pairs of codewords of different blocklengths are juxtaposed. Say, $(\underline{x}_1, \underline{x}_2)$ has blocklength n while $(\underline{x}'_1, \underline{x}'_2)$ has blocklength n'. Then

$$\tau_{\widetilde{\underline{x}}_1,\widetilde{\underline{x}}_2} = \frac{n}{n+n'}\tau_{\underline{x}_1,\underline{x}_2} + \frac{n'}{n+n'}\tau_{\underline{x}'_1,\underline{x}'_2}.$$

Back to the proof of the converse in such a spacial case, since the confusability set is defined by a single linear constraint, any convex combinations of non-confusable joint types is still outside the confusability set, in particular, $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$. We hence reduce the problem to the symmetric case and the rest of the proof is handled by Theorem 65. ◀

14.2 Towards a unifying converse

We feel it unusual that we have to use drastically different techniques to prove the symmetric and the asymmetric parts of the converse. We suspect that it can be proved in a unifying way using the duality between CP and coP tensors which is the source of contradiction in our current proof of the symmetric case.

Note that the duality holds only in the space of *symmetric* tensors. To be specific, traditionally, CP and coP tensors are defined to be symmetric. And they are dual cones living in the ambient space $\mathsf{Sym}_n^{\otimes}$. If we extend the definitions of CP and coP tensors to the set of *all* (including asymmetric) tensors, then it is unclear whether duality still holds. Indeed, there are pairs of cones which are dual to each other in a certain ambient space but are no long dual in a larger ambient space. In a word, the ambient space that the dual cone is computed with respect to matters much.

We provide evidence showing that the symmetric and asymmetric parts of the converse can be potentially unified by the Plotkin-type bound since duality between CP and coP tensors–the core of the double counting argument–fortunately holds in larger generality.

Duality. We know that $\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}$ and $\mathsf{coP}^{\otimes L}_{|\mathcal{X}|}$ are dual cones in the space $\mathsf{Sym}^{\otimes L}_{|\mathcal{X}|}$ of *symmetric* tensors. However, $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ (associated to the equicoupled subcode extracted using hypergraph Ramsey's theorem) is not guaranteed to be symmetric. We claim that duality still holds in the space $\mathsf{Ten}^{\otimes L}_{|\mathcal{X}|}$ of *all* tensors. Hence, copositive witness Q of a non-CP $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ exists even when $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is asymmetric.

▷ Claim 75. $\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}$ and $\mathsf{coP}^{\otimes L}_{|\mathcal{X}|}$ are dual cones in $\mathsf{Ten}^{\otimes L}_{|\mathcal{X}|}$.

Proof. By definition,

$$\left(\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}\right)^* := \left\{B \in \mathsf{Ten}^{\otimes L}_{|\mathcal{X}|} \colon \forall A \in \mathsf{CP}^{\otimes L}_{|\mathcal{X}|},\, \langle A, B\rangle \geqslant 0\right\}.$$

Note that it is important that B is now taken from $\mathsf{Ten}^{\otimes L}_{|\mathcal{X}|}$ rather than $\mathsf{Sym}^{\otimes L}_{|\mathcal{X}|}$. Also recall that

$$\mathsf{coP}^{\otimes L}_{|\mathcal{X}|} := \left\{B \in \mathsf{Ten}^{\otimes L}_{|\mathcal{X}|} \colon \forall \underline{x} \in \mathbb{R}^{|\mathcal{X}|}_{\geqslant 0},\, \left\langle B, \underline{x}^{\otimes L}\right\rangle \geqslant 0\right\}.$$

Note that this definition *differs* from the standard one 95 and this cone is potentially larger.[16] The goal is to show $\left(\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}\right)^* = \mathsf{coP}^{\otimes L}_{|\mathcal{X}|}$.

The direction $\mathsf{coP}_{|\mathcal{X}|} \subseteq \left(\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}\right)^*$ is trivial, since the definitions of CP and coP tensors remain the same but the dual cone is computed w.r.t. a larger space. The new dual cone we are considering is no smaller than the old one. The inclusion that used to hold in the traditional setting should continue to hold now. Indeed, take any $B \in \mathsf{coP}^{\otimes L}_{|\mathcal{X}|}$, for any $A = \sum_i \underline{x}_i^{\otimes L} \in \mathsf{CP}^{\otimes L}_{|\mathcal{X}|}$, where $\underline{x}_i \in \mathbb{R}^{|\mathcal{X}|}_{\geqslant 0}$,

$$\langle A, B\rangle = \left\langle \sum_i \underline{x}_i^{\otimes L}, B \right\rangle = \sum_i \left\langle B, \underline{x}_i^{\otimes L}\right\rangle.$$

Since $B \in \mathsf{coP}^{\otimes L}_{|\mathcal{X}|}$, by definition, all $\left\langle B, \underline{x}_i^{\otimes L}\right\rangle$'s are non-negative, hence so is $\langle A, B\rangle$. Therefore $B \in \left(\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}\right)^*$.

Now we show $\left(\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}\right)^* \subseteq \mathsf{coP}^{\otimes L}_{|\mathcal{X}|}$. Take any $B \in \left(\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}\right)^*$ and any $\underline{x} \in \mathbb{R}^{|\mathcal{X}|}_{\geqslant 0}$. Then $\left\langle B, \underline{x}^{\otimes L}\right\rangle \geqslant 0$, since $\underline{x}^{\otimes L} \in \mathsf{CP}^{\otimes L}_{|\mathcal{X}|}$ and $B \in \left(\mathsf{CP}^{\otimes L}_{|\mathcal{X}|}\right)^*$. This finishes the whole proof. ◁

▶ Remark 76. In general, duality does not necessarily hold in a larger ambient space. Namely, computing dual cone w.r.t. a larger space may result in a larger cone. For instance, $\mathsf{PSD}_{|\mathcal{X}|}$ cone is known to be self dual in $\mathsf{Sym}_{|\mathcal{X}|}$, i.e., $\mathsf{PSD}^*_{|\mathcal{X}|} = \mathsf{PSD}_{|\mathcal{X}|}$. However, in $\mathsf{Mat}_{|\mathcal{X}|}$, $\mathsf{PSD}^*_{|\mathcal{X}|}$ is strictly containing $\mathsf{PSD}_{|\mathcal{X}|}$. To see this, note that any skew symmetric matrix B is in $\mathsf{PSD}^*_{|\mathcal{X}|}$ since for any PSD (hence symmetric) matrix A, $\langle A, B\rangle = 0 \geqslant 0$; while B is not necessarily PSD.

Define, for $\sigma \in S_L$, $\sigma\left(P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}\right) := P_{\mathbf{x}_{\sigma(1)},\cdots,\mathbf{x}_{\sigma(L)}}$. Though duality holds for all symmetric and asymmetric tensors, we do not have a full proof of the converse using duality, since we have trouble bounding the term

$$\left\langle \sigma\left(P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}\right), Q\right\rangle = \left\langle P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}, \sigma(Q)\right\rangle$$

which does not necessarily equal $\langle P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}, Q\rangle$ for *asymmetric* Q.

We next show that such asymmetric witness Q does exist and is sometimes necessary in the sense that, some asymmetric (hence non-CP) tensors have *no* symmetric witness. This means that the dual cone of coP w.r.t. $\mathsf{Ten}^{\otimes L}_{|\mathcal{X}|}$ (instead of $\mathsf{Sym}^{\otimes L}_{|\mathcal{X}|}$) is strictly larger.

[16] Indeed, we will see shortly that it is strictly larger.

Asymmetric distributions without symmetric coP witness. Let $L = 2$. We construct an asymmetric self-coupling $P_{\mathbf{x}_1,\mathbf{x}_2} \in \Delta\left([3]^2\right)$ without symmetric coP witness Q such that $\langle P_{\mathbf{x}_1,\mathbf{x}_2}, Q\rangle < 0$. Indeed, let

$$P_{\mathbf{x}_1,\mathbf{x}_2} = \begin{bmatrix} \frac{4}{9} & \frac{7}{48} & \frac{11}{144} \\ \frac{3}{16} & \frac{1}{16} & 0 \\ \frac{5}{144} & \frac{1}{24} & \frac{1}{144} \end{bmatrix}.$$

Note that

$$P_{\mathbf{x}_1} = P_{\mathbf{x}_2} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{4} \\ \frac{1}{12} \end{bmatrix} =: P_{\mathbf{x}}.$$

Then

$$\frac{P_{\mathbf{x}_1,\mathbf{x}_2} + P_{\mathbf{x}_1,\mathbf{x}_2}^\top}{2} = \begin{bmatrix} \frac{4}{9} & \frac{1}{6} & \frac{1}{18} \\ \frac{1}{6} & \frac{1}{16} & \frac{1}{48} \\ \frac{1}{18} & \frac{1}{48} & \frac{1}{144} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{4} \\ \frac{1}{12} \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{1}{4} & \frac{1}{12} \end{bmatrix} = P_{\mathbf{x}} P_{\mathbf{x}}^\top.$$

If there was a symmetric coP Q such that $\langle P_{\mathbf{x}_1,\mathbf{x}_2}, Q\rangle < 0$, then

$$\begin{aligned} \left\langle P_{\mathbf{x}} P_{\mathbf{x}}^\top, Q\right\rangle &= \frac{1}{2}\left(\langle P_{\mathbf{x}_1,\mathbf{x}_2}, Q\rangle + \left\langle P_{\mathbf{x}_1,\mathbf{x}_2}^\top, Q\right\rangle\right) \\ &= \frac{1}{2}\left(\langle P_{\mathbf{x}_1,\mathbf{x}_2}, Q\rangle + \left\langle P_{\mathbf{x}_1,\mathbf{x}_2}, Q^\top\right\rangle\right) \\ &= \langle P_{\mathbf{x}_1,\mathbf{x}_2}, Q\rangle < 0. \end{aligned}$$

However, $P_{\mathbf{x}} P_{\mathbf{x}}^\top$ is CP, so $\left\langle P_{\mathbf{x}} P_{\mathbf{x}}^\top, Q\right\rangle \geqslant 0$, which is a contradiction.

15 Sanity checks

Consider the bit-flip model.

In this section, we are going to verify the correctness of our characterization of the generalized Plotkin point using the bit-flip model as a running example. For $L = 3, 4$,[17] we will numerically recover Blinovsky's [9] characterization of the Plotkin point P_{L-1} for $(p, L-1)$-list decoding. In particular, $P_2 = 1/4$ and $P_3 = 5/16$.

15.1 $L = 3$

We first consider $(L-1)$-list decoding for $L - 1 = 2$, i.e., $L = 3$. It is known that the Plotkin point at $L - 1 = 2$ is $P_2 = 1/4$.

Fix any input distribution $P_{\mathbf{x}} := \mathrm{Bern}(w) = \begin{bmatrix} 1-w \\ w \end{bmatrix}$ for $0 < w < 1$. We first compute $\mathcal{J}^{\otimes 3}(P_{\mathbf{x}})$, $\mathcal{K}^{\otimes 3}(P_{\mathbf{x}})$. Let $p_{i,j,k,\ell} := P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}}(i,j,k,\ell)$ where $i,j,k,\ell \in \{0,1\}$.

$$\mathcal{J}^{\otimes 3}(P_{\mathbf{x}}) = \left\{P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \in \Delta\left(\{0,1\}^3\right) : P_{\mathbf{x}_i} = P_{\mathbf{x}},\ i = 1,2,3\right\}$$

[17] For $L = 2$, i.e., the unique decoding case, the work [43] already recovers the classic Plotkin bound $P_1 = 1/4$.

$$= \left\{ P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} : \begin{array}{rl} p_{i,j,k} \geqslant & 0,\ i,j,k \in \{0,1\} \\ \sum_{i,j,k} p_{i,j,k} = & 1 \\ \sum_{i,j} p_{i,j,1} = & w \\ \sum_{i,k} p_{i,1,k} = & w \\ \sum_{j,k} p_{1,j,k} = & w \end{array} \right\}.$$

$$\begin{aligned} &\mathcal{K}^{\otimes 3}(P_{\mathbf{x}}) \\ &= \left\{ P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} = [P_{\mathbf{x}_2,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}}]_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \in \mathcal{J}^{\otimes 3}(P_{\mathbf{x}}) : \begin{array}{rl} P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}} \in & \Delta\left(\{0,1\}^4\right) \\ P_{\mathbf{x}_i,\mathbf{y}}(0,1) + P_{\mathbf{x}_i,\mathbf{y}}(1,0) \leqslant & p,\ i = 1,2,3 \end{array} \right\} \\ &= \left\{ [P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}}]_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} \in \mathcal{J}^{\otimes 3}(P_{\mathbf{x}}) : \begin{array}{rl} p_{i,j,k,\ell} \geqslant & 0,\ i,j,k,\ell \in \{0,1\} \\ \sum_{i,j,k,\ell} p_{i,j,k,\ell} = & 1 \\ \sum_{j,k} p_{0,j,k,1} + p_{1,j,k,0} \leqslant & p \\ \sum_{i,k} p_{i,0,k,1} + p_{i,1,k,0} \leqslant & p \\ \sum_{i,j} p_{i,j,0,1} + p_{i,j,1,0} \leqslant & p \end{array} \right\}. \end{aligned}$$

$\widehat{\mathcal{J}}^{\otimes(L+1)}(P_{\mathbf{x}})$ and $\widehat{\mathcal{K}}^{\otimes(L+1)}(P_{\mathbf{x}})$ are extended formulations of $\mathcal{J}^{\otimes L}(P_{\mathbf{x}})$ and $\mathcal{K}^{\otimes L}(P_{\mathbf{x}})$, respectively.

$$\widehat{\mathcal{J}}^{\otimes 4}(P_{\mathbf{x}}) = \left\{ P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}} : \begin{array}{rl} p_{i,j,k,\ell} \geqslant & 0,\ i,j,k,\ell \in \{0,1\} \\ \sum_{i,j,k,\ell \in \{0,1\}} p_{i,j,k,\ell} = & 1 \\ \sum_{i,j} p_{i,j,1} = & w \\ \sum_{i,k} p_{i,1,k} = & w \\ \sum_{j,k} p_{1,j,k} = & w \end{array} \right\}.$$

$$\widehat{\mathcal{K}}^{\otimes 4}(P_{\mathbf{x}}) = \left\{ P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}} \in \widehat{\mathcal{J}}^{\otimes 4}(P_{\mathbf{x}}) : \begin{array}{rl} \sum_{j,k \in \{0,1\}} p_{0,j,k,1} + p_{1,j,k,0} \leqslant & p \\ \sum_{i,k \in \{0,1\}} p_{i,0,k,1} + p_{i,1,k,0} \leqslant & p \\ \sum_{i,j \in \{0,1\}} p_{i,j,0,1} + p_{i,j,1,0} \leqslant & p \end{array} \right\}.$$

To verify the value of Plotkin point P_{L-1} at $L = 3$, it suffices to verify that, if $w = 1/2$, then $P_{\mathbf{x}}^{\otimes 3} \notin \mathcal{K}^{\otimes 3}(P_{\mathbf{x}})$ iff $p < 1/4$, since we know that the optimizing input distribution when codewords are weight unconstrained is uniform. To this end, define a hyperplane

$$\mathcal{H}\left(P_{\mathbf{x}}^{\otimes 3}\right) := \left\{ P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}} \in \widehat{\mathcal{J}}^{\otimes 4}(P_{\mathbf{x}}) : [P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}}]_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} = P_{\mathbf{x}}^{\otimes 3} \right\}.$$

Note that $P_{\mathbf{x}}^{\otimes 3} \notin \mathcal{K}^{\otimes 3}(P_{\mathbf{x}})$ is equivalent to $\mathcal{H}\left(P_{\mathbf{x}}^{\otimes 3}\right) \cap \widehat{\mathcal{K}}^{\otimes 4}(P_{\mathbf{x}}) = \varnothing$. Since $\mathcal{H}\left(P_{\mathbf{x}}^{\otimes 3}\right)$ depends on w and $\widehat{\mathcal{K}}^{\otimes 4}(P_{\mathbf{x}})$ depends on w, p, we write them as $\mathcal{H}(w)$ and $\widehat{\mathcal{K}}^{\otimes 4}(w,p)$, respectively, for simplicity.

We claim that the Plotkin point P_{L-1} is precisely the optimal value of the following LP, i.e., the smallest p^* such that the hyperplane $\mathcal{H}(1/2)$ has no intersection with the corresponding high-dimensional polytope $\widehat{\mathcal{K}}^{\otimes 4}(1/2, p^*)$.

$$\begin{array}{rl} \min & p \\ \text{subject to} & \mathcal{H}(1/2) \cap \widehat{\mathcal{K}}^{\otimes 4}(1/2,p) \neq \varnothing. \end{array}$$

Equivalently, collecting all constraints together, we want to find the minimal p so that the polytope (the feasible region of the LP) defined by the following constraints is nonempty.

$$P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}} \in \widehat{\mathcal{J}}^{\otimes 4}(P_{\mathbf{x}})$$

$$
\begin{aligned}
[P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}}]_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} =& P_{\mathbf{x}}^{\otimes 3}\\
\sum_{j,k\in\{0,1\}} p_{0,j,k,1} + p_{1,j,k,0} \leqslant& p\\
\sum_{i,k\in\{0,1\}} p_{i,0,k,1} + p_{i,1,k,0} \leqslant& p\\
\sum_{i,j\in\{0,1\}} p_{i,j,0,1} + p_{i,j,1,0} \leqslant& p.
\end{aligned}
$$

Expanding everything out and noting that the first constraint regarding constant composition $P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}\in\hat{\mathcal{J}}^{\otimes 4}(P_{\mathbf{x}})}$ is redundant since it is the same as the constraint $[P_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3,\mathbf{y}}]_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}_3} = P_{\mathbf{x}}^{\otimes 3} \in \mathcal{J}^{\otimes 3}(P_{\mathbf{x}})$, we simplify the defining (in)equalities of the polytope as follows,

$$
\begin{aligned}
p_{i,j,k,\ell} \geqslant& 0,\ i,j,k,\ell \in \{0,1\}\\
\sum_{i,j,k,\ell\in\{0,1\}} p_{i,j,k,\ell} =& 1\\
p_{i,j,k,0} + p_{i,j,k,1} =& 1/8,\ i,j,k \in \{0,1\}\\
\sum_{j,k\in\{0,1\}} p_{0,j,k,1} + p_{1,j,k,0} \leqslant& p\\
\sum_{i,k\in\{0,1\}} p_{i,0,k,1} + p_{i,1,k,0} \leqslant& p\\
\sum_{i,j\in\{0,1\}} p_{i,j,0,1} + p_{i,j,1,0} \leqslant& p,
\end{aligned}
$$

since $P_{\mathbf{x}}^{\otimes 3}(i,j,k) = P_{\mathbf{x}}(i)P_{\mathbf{x}}(j)P_{\mathbf{x}}(k) = 1/8$ for all $i,j,k \in \{0,1\}$.

Let

$$\underline{p} \coloneqq \begin{bmatrix} p_{0,0,0,0} & \cdots & p_{1,1,1,1} \end{bmatrix}^\top.$$

The LP can be written in a compact form as

$$
\begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\
1 & 1 & & & & & & & & & & & & & & \\
 & & 1 & 1 & & & & & & & & & & & & \\
 & & & & 1 & 1 & & & & & & & & & & \\
 & & & & & & 1 & 1 & & & & & & & & \\
 & & & & & & & & 1 & 1 & & & & & & \\
 & & & & & & & & & & 1 & 1 & & & & \\
 & & & & & & & & & & & & 1 & 1 & & \\
 & & & & & & & & & & & & & & 1 & 1
\end{bmatrix}
\underline{p} =
\begin{bmatrix} 1\\1/8\\1/8\\1/8\\1/8\\1/8\\1/8\\1/8\\1/8 \end{bmatrix},
$$

$$
\begin{bmatrix}
1 & & 1 & & 1 & & 1 & 1 & & 1 & & 1 & & 1 & \\
1 & & 1 & 1 & & 1 & & & 1 & & 1 & 1 & & 1 & \\
1 & 1 & & & 1 & 1 & & & 1 & 1 & & & 1 & 1 &
\end{bmatrix}
\underline{p} \leqslant
\begin{bmatrix} p\\p\\p \end{bmatrix}
$$

$$\underline{p} \geqslant \underline{0}.$$

Observe that as p increases, the linear system becomes monotonically easier to be satisfied. Checked by `Mathematica`, the above LP is feasible if $p > 1/4$ (and hence the distribution $\begin{bmatrix}1/2\\1/2\end{bmatrix}^{\otimes 3}$ is confusable) and is infeasible if $p < 1/4$ (and hence $\begin{bmatrix}1/2\\1/2\end{bmatrix}^{\otimes 3}$ is not confusable). Therefore, the $(p, L-1)$-list decoding capacity hits 0 precisely at $p = 1/4$.

15.2 $L = 4$

For $L = 4$, one can obtain a similar LP whose infeasibility is equivalent to $\mathcal{H}\left(\begin{bmatrix}1/2\\1/2\end{bmatrix}^{\otimes 4}\right)$ and $\widehat{\mathcal{K}}^{\otimes 5}\left(\begin{bmatrix}1/2\\1/2\end{bmatrix}^{\otimes 4}, p\right)$ being disjoint.

$$\begin{bmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&&&&&&&&&&&&&&&&&&&&&&&&&&&&\\
&&1&1&&&&&&&&&&&&&&&&&&&&&&&&&&\\
&&&&1&1&&&&&&&&&&&&&&&&&&&&&&&&\\
&&&&&&1&1&&&&&&&&&&&&&&&&&&&&&&\\
&&&&&&&&1&1&&&&&&&&&&&&&&&&&&&&\\
&&&&&&&&&&1&1&&&&&&&&&&&&&&&&&&\\
&&&&&&&&&&&&1&1&&&&&&&&&&&&&&&&\\
&&&&&&&&&&&&&&1&1&&&&&&&&&&&&&&\\
&&&&&&&&&&&&&&&&1&1&&&&&&&&&&&&\\
&&&&&&&&&&&&&&&&&&1&1&&&&&&&&&&\\
&&&&&&&&&&&&&&&&&&&&1&1&&&&&&&&\\
&&&&&&&&&&&&&&&&&&&&&&1&1&&&&&&\\
&&&&&&&&&&&&&&&&&&&&&&&&1&1&&&&\\
&&&&&&&&&&&&&&&&&&&&&&&&&&1&1&&\\
&&&&&&&&&&&&&&&&&&&&&&&&&&&&1&1
\end{bmatrix}\underline{p} = \begin{bmatrix}1\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\\1/16\end{bmatrix},$$

$$\begin{bmatrix}
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&1&1&1&1&1&1&&&&&&&&\\
1&1&1&1&&&&&1&1&1&1&&&&\\
1&1&&&1&1&&&1&1&&&1&1&&\\
&&1&1&&&1&1&&&1&1&&&1&1
\end{bmatrix}\underline{p} \leqslant \begin{bmatrix}p\\p\\p\\p\end{bmatrix},$$

$$\underline{p} \geqslant \underline{0},$$

where

$$\underline{p} = \begin{bmatrix}p_{0,0,0,0,0} & \cdots & p_{1,1,1,1,1}\end{bmatrix}^{\top}.$$

One can numerically check that the above LP is feasible if $p > 5/16$ and infeasible otherwise.

In general, to check whether

$$\mathcal{H}\left(\begin{bmatrix}1/2\\1/2\end{bmatrix}^{\otimes L}\right) \cap \widehat{\mathcal{K}}^{\otimes(L+1)}\left(\begin{bmatrix}1/2\\1/2\end{bmatrix}^{\otimes L}, p\right)$$

is empty, it boils down to checking the infeasibility of a linear program with 2^{L+1} variables and $2^{L+1} + 1 + 2^L + L$ constraints, 2^{L+1} of them for non-negativity of probability mass, 1 of them for probability mass summing up to one, 2^L of them for ensuring that $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \mathcal{J}^{\otimes L}(P_{\mathbf{x}})$ is a $(P_{\mathbf{x}}, L)$-self-coupling, L of them for the non-confusability guarantee: $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \notin \mathcal{K}^{\otimes L}(P_{\mathbf{x}})$. The size of the program (or the number of defining constraints of the corresponding polytope) grows exponentially in L. However, since we are concerned with absolute constant L in this paper, for any given L, the feasibility can be certified in constant time. Observe that, since the LP in the bit-flip setting is so structured, one can write it down explicitly by hand for any given L.

16 Blinovsky [9] revisited

In this section, we *fully* recover Blinovsky's [9] results on characterization of the Plotkin points P_{L-1} for $(p, L-1)$-list decoding under the bit-flip model.

Let ϕ be the standard bijection between $\{0,1\}$ and $\{-1,1\}$,

$$\begin{array}{rccc}
\phi\colon & \mathbb{F}_2 & \to & \{-1,1\}\\
& 0 & \mapsto & 1\\
& 1 & \mapsto & -1.
\end{array}$$

We identify the type $\tau_{\underline{x}} \in \mathcal{P}^{(n)}(\mathbb{F}_2)$ of a binary length-n vector $\underline{x} \in \mathbb{F}_2^n$ using a $\{-1,1\}$-valued random variable $\mathbf{x}$ defined as

$$\Pr[\mathbf{x}=-1] = \frac{wt_{\mathrm{H}}(\underline{x})}{n}, \quad \Pr[\mathbf{x}=1] = 1 - \frac{wt_{\mathrm{H}}(\underline{x})}{n}.$$

Indeed the distribution $P_{\mathbf{x}} \in \mathcal{P}^{(n)}(\{-1,1\})$ of $\mathbf{x}$ is the type of the image $\phi(\underline{x})$ of $\underline{x}$ under ϕ.

$$P_{\mathbf{x}}(\phi(0)) = \tau_{\underline{x}}(0), \quad P_{\mathbf{x}}(\phi(1)) = \tau_{\mathbf{x}}(1).$$

For a collection of vectors $\underline{x}_1, \cdots, \underline{x}_k \in \mathbb{F}_2^n$, their joint type is now represented by a sequence of random variables $\mathbf{x}_1, \cdots, \mathbf{x}_k$ with joint distribution $P_{\mathbf{x}_1,\cdots,\mathbf{x}_k}$, for any $x_1, \cdots, x_k \in \{-1,1\}$,

$$\begin{aligned} P_{\mathbf{x}_1,\cdots,\mathbf{x}_k}(x_1,\cdots,x_k) =& \Pr[\mathbf{x}_1 = x_1, \cdots, \mathbf{x}_k = x_k] \\ =& \tau_{\underline{x}_1,\cdots,\underline{x}_k}(\phi^{-1}(x_1), \cdots, \phi^{-1}(x_k)). \end{aligned}$$

It is easy to check that, for $\underline{x}_1, \underline{x}_2 \in \mathbb{F}_2^n$,

$$\frac{d_{\mathrm{H}}(\underline{x}_1, \underline{x}_2)}{n} = \frac{1}{2}\left(1 - \mathop{\mathbb{E}}_{(\mathbf{x}_1,\mathbf{x}_2)\sim P_{\mathbf{x}_1,\mathbf{x}_2}}[\mathbf{x}_1\mathbf{x}_2]\right). \tag{102}$$

Indeed

$$\begin{aligned} \text{RHS} =& \frac{1}{2}\left(1 - \tau_{\underline{x}_1,\underline{x}_2}(0,1)\cdot(-1) - \tau_{\underline{x}_1,\underline{x}_2}(1,0)\cdot(-1) - \tau_{\underline{x}_1,\underline{x}_2}(0,0)\cdot 1 - \tau_{\underline{x}_1,\underline{x}_2}(1,1)\cdot 1\right) \\ =& \frac{1}{2}\left(1 + \frac{d_{\mathrm{H}}(\underline{x}_1,\underline{x}_2)}{n} - \left(1 - \frac{d_{\mathrm{H}}(\underline{x}_1,\underline{x}_2)}{n}\right)\right) \\ =& \text{LHS}. \end{aligned}$$

Let

$$r := \mathop{\mathbb{E}}_{(\mathbf{x}_1,\cdots,\mathbf{x}_L)\sim\{-1,1\}^L}\left[|\mathbf{x}_1 + \cdots + \mathbf{x}_L|\right], \tag{103}$$

be the expected translation distance of a 1-dimensional unbiased random walk after L steps. Each $\mathbf{x}_i$ $(1 \leqslant i \leqslant L)$ is independent and uniformly distributed on $\{-1,1\}$.

▶ **Theorem 77.** *The Plotkin point P_{L-1} for $(p, L-1)$-list decoding is given by*

$$P_{L-1} = \frac{1 - r/L}{2}.$$

▶ Remark 78. Note that the formula in Theorem 77 agrees with the one by Blinovsky. To see this, we first compute r. For odd $L = 2k+1$, where $k \in \mathbb{Z}_{>0}$ is some strictly positive integer, it is easy to see that

$$\begin{aligned} r =& \mathbb{E}\left[|\mathbf{x}_1 + \cdots + \mathbf{x}_L|\right] \\ =& \sum_{i=0}^{k} \frac{2\binom{L}{i}}{2^L}(L-2i). \end{aligned}$$

Recall that, by binomial theorem (Fact (18)),

$$2^L = \sum_{i=0}^{L}\binom{L}{i} = \sum_{i=0}^{k} 2\binom{L}{i}.$$

Now we simplify the formula in Theorem 77.

$$\begin{aligned}
P_{L-1} =& \frac{1}{2} - \frac{r}{2L} \\
=& \sum_{i=0}^{k} \frac{\binom{L}{i}}{2^L} - \sum_{i=0}^{k} \left(1 - \frac{2i}{L}\right) \frac{\binom{L}{i}}{2^L} \\
=& \sum_{i=0}^{k} \frac{2i}{L} \frac{\binom{L}{i}}{2^L} \\
=& \sum_{i=1}^{k} \frac{i}{L} \frac{\frac{L}{i}\binom{L-1}{i-1}}{2^{L-1}} && (104) \\
=& \frac{1}{2^{L-1}} \sum_{i=0}^{k-1} \binom{L-1}{i} \\
=& \frac{1}{2^{L-1}} \frac{1}{2} \left(2^{L-1} - \binom{L-1}{k}\right) && (105) \\
=& \frac{1}{2} - 2^{-L} \binom{2k}{k},
\end{aligned}$$

where Eqn. (104) is by Fact (16); Eqn. (105) follows from binomial theorem (Fact (18)) again,

$$2^{L-1} = \binom{2k}{k} + 2 \sum_{i=0}^{k-1} \binom{2k}{i}.$$

▶ **Lemma 79** (Lower bound). *The Plotkin point P_{L-1} for $(p, L-1)$-list decoding is lower bounded by*

$$P_{L-1} \geqslant \frac{1 - r/L}{2}.$$

That is, if $p < P_{L-1}$, then the $(p, L-1)$-list decoding capacity is positive, i.e., there is an infinite sequence of $(p, L-1)$-list decodable codes of positive rate.

Proof. We will show that if $p = \frac{1 - \frac{r+\eta}{L}}{2} < \frac{1-r/L}{2}$ for any $\eta > 0$, then the product distribution $\mathrm{Bern}^{\otimes L}(1/2)$ lies outside the corresponding confusability set $\mathcal{K}^{\otimes L}(\mathrm{Bern}(1/2))$. Using the framework developed in this paper, a random code of a suitable positive rate in which each codeword is sampled independently and uniformly from $\mathcal{T}_{\underline{\mathbf{x}}}(\mathrm{Bern}(1/2))$ is $(p, L-1)$-list decodable w.h.p.

The proof is by contradiction. If $P_{\mathbf{x}_1,\cdots,\mathbf{x}_L} := \mathrm{Bern}^{\otimes L}(1/2)$ is confusable, then, by the definition 37 of confusability of tuples, an L-tuple of distinct codewords $\underline{x}_1, \cdots, \underline{x}_L$ of joint type $\tau_{\underline{x}_1,\cdots,\underline{x}_L} = P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ can be covered by a ball of radius np centered around some $\underline{y} \in \mathbb{F}_2^n$. Equivalently, by the definition 38 of confusability of distributions, there is a refinement $P_{\mathbf{x}_1,\cdots,\mathbf{x},\mathbf{y}} \in \Delta\left(\{-1,1\}^{L+1}\right)$ such that $[P_{\mathbf{x}_1,\cdots,\mathbf{x}_L,\mathbf{y}}]_{\mathbf{x}_1,\cdots,\mathbf{x}_L} = P_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$, and for every $i \in [L]$,

$$P_{\mathbf{x}_i,\mathbf{y}}(0,1) + P_{\mathbf{x}_i,\mathbf{y}}(1,0) \leqslant p.$$

This means that for every $i \in [L]$,

$$\mathbb{E}[\mathbf{x}_i \mathbf{y}] \geqslant \frac{r+\eta}{L},$$

by the relation (Eqn. (102)) between Hamming distance between vectors and correlation of their random variable representations. Hence

$$\mathbb{E}\left[(\mathbf{x}_1 + \cdots + \mathbf{x}_L)\,\mathbf{y}\right] \geqslant r + \eta. \tag{106}$$

The $\{-1,1\}$-valued random variable $\mathbf{y}$ that has the largest correlation with $\mathbf{x}_1 + \cdots + \mathbf{x}_L$ is $\mathbf{y} = \mathsf{MAJ}\left(\mathbf{x}_1, \cdots, \mathbf{x}_L\right)$, where

$$\begin{array}{rccc} \mathsf{MAJ}: & \{-1,1\}^L & \to & \{-1,1\} \\ & (x_1, \cdots, x_L) & \mapsto & \operatorname{sgn}\left(x_1 + \cdots + x_L\right). \end{array}$$

is the majority function. To see this, just expand the above expectation,

$$\begin{aligned} \mathbb{E}\left[(\mathbf{x}_1 + \cdots + \mathbf{x}_L)\,\mathbf{y}\right] &= \sum_{x_1, \cdots, x_L, y \in \{-1,1\}} P_{\mathbf{x}_1, \cdots, \mathbf{x}_L, \mathbf{y}}(x_1, \cdots, x_L, y)(x_1 + \cdots + x_L)y \\ &= \sum_{x_1, \cdots, x_L \in \{-1,1\}} P_{\mathbf{x}_1, \cdots, \mathbf{x}_L}(x_1, \cdots, x_L) \\ &\qquad \sum_{y \in \{-1,1\}} P_{\mathbf{y}|\mathbf{x}_1, \cdots, \mathbf{x}_L}(y|x_1, \cdots, x_L)(x_1 + \cdots + x_L)y. \end{aligned}$$

Note that, each summand

$$P_{\mathbf{y}|\mathbf{x}_1, \cdots, \mathbf{x}_L}(1|x_1, \cdots, x_L)(x_1 + \cdots + x_L) - P_{\mathbf{y}|\mathbf{x}_1, \cdots, \mathbf{x}_L}(-1|x_1, \cdots, x_L)(x_1 + \cdots + x_L)$$

is maximized when the conditional probability mass of $\mathbf{y}$ is concentrated on the singleton $\operatorname{sgn}(x_1 + \cdots + x_L)$,

$$P_{\mathbf{y}|\mathbf{x}_1, \cdots, \mathbf{x}_L}(\operatorname{sgn}(x_1 + \cdots + x_L)|x_1, \cdots, x_L) = 1, P_{\mathbf{y}|\mathbf{x}_1, \cdots, \mathbf{x}_L}(-\operatorname{sgn}(x_1 + \cdots + x_L)|x_1, \cdots, x_L) = 0.$$

In this case, each summand attains its maxima

$$\operatorname{sgn}(x_1 + \cdots + x_L)(x_1 + \cdots + x_L) = |x_1 + \cdots + x_L|.$$

Overall, the corresponding maximal correlation is precisely

$$\begin{aligned} &\mathbb{E}(\mathbf{x}_1 + \cdots + \mathbf{x}_L)\mathsf{MAJ}(\mathbf{x}_1, \cdots, \mathbf{x}_L) \\ &= \sum_{x_1, \cdots, x_L \in \{-1,1\}} P_{\mathbf{x}_1, \cdots, \mathbf{x}_L}(x_1, \cdots, x_L)\,|x_1 + \cdots + x_L| \\ &= \mathop{\mathbb{E}}_{(\mathbf{x}_1, \cdots, \mathbf{x}_L) \sim P_{\mathbf{x}_1, \cdots, \mathbf{x}_L}} \left[|\mathbf{x}_1 + \cdots + \mathbf{x}_L|\right]. \end{aligned} \tag{107}$$

Using the above observation, we get

$$\begin{aligned} r &= \mathop{\mathbb{E}}_{(\mathbf{x}_1, \cdots, \mathbf{x}_L) \sim \{-1,1\}^L} \left[|\mathbf{x}_1 + \cdots + \mathbf{x}_L|\right] && (108) \\ &= \mathbb{E}\left[(\mathbf{x}_1 + \cdots + \mathbf{x}_L)\,\mathsf{MAJ}\left(\mathbf{x}_1, \cdots, \mathbf{x}_L\right)\right] && (109) \\ &\geqslant r + \eta, && (110) \end{aligned}$$

Eqn. (108) is by the definition of r (Eqn. (103)). Eqn. (109) follows from Eqn. (107). Eqn. (110) is by Eqn. (110). We hence reach a contradiction which finishes the proof. ◀

▶ **Lemma 80** (Upper bound). *The Plotkin point P_{L-1} for $(p, L-1)$-list decoding is upper bounded by*

$$P_{L-1} \leqslant \frac{1 - r/L}{2}.$$

That is, if $p > P_{L-1}$, then no positive rate is possible, i.e, there is no infinite sequence of $(p, L-1)$-list decodable codes of positive rate.

Proof. Our goal is to show that if $p > P_{L-1}$, then $C_{L-1} = 0$. Suppose $p = \frac{1-\frac{r-\eta}{L}}{2}$ for a constant $\eta > 0$.

We are going to show that any infinite sequence of codes $\mathcal{C}_n$ each of positive rate is not $(p, L-1)$-list decodable. First, by the previous argument in last section, we can extract a sequence of subcodes $\mathcal{C}'_n \subseteq \mathcal{C}_n$ of positive rate satisfying that, for every tuple of *distinct* codewords $\underline{x}_1, \cdots, \underline{x}_L \in \mathcal{C}'$ and $x_1, \cdots, x_L \in \mathbb{F}_2$,

$$\left|\tau_{\underline{x}_1,\cdots,\underline{x}_L}(x_1,\cdots,x_L) - \widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(x_1,\cdots,x_L)\right| \leqslant \zeta$$

for some *symmetric* distribution $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L} \in \Delta\left(\mathcal{X}^L\right)$ and some positive constant $\zeta > 0$. In favour of the proceeding calculations, it suffices to take

$$\zeta = \frac{L}{(L-1)r2^{L+2}}\eta. \tag{111}$$

To show non-list decodability of $\mathcal{C}'$ (and hence $\mathcal{C}$), we will argue that there is a list $(\underline{x}_{i_1}, \cdots, \underline{x}_{i_L}) \in \binom{\mathcal{C}'}{L}$ that can be covered by a ball of radius np centered around the point $\mathsf{MAJ}\left(\underline{x}_{i_1}, \cdots, \underline{x}_{i_L}\right)$. The proof is by contradiction. Suppose this is not the case, i.e., no list can be covered by the ball centered at its majority. Define, for $(i_1, \cdots, i_L) \in \left[2^{nR}\right]^L$,

$$Q_{i_1,\cdots,i_L} = (\mathbf{x}_{i_1} + \cdots + \mathbf{x}_{i_L}) \cdot \mathsf{MAJ}\left(\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_L}\right) - r.$$

We will provide a *strictly negative* upper bound and a *non-negative* lower bound on

$$Q \coloneqq \mathop{\mathbb{E}}_{(\mathbf{i}_1,\cdots,\mathbf{i}_L)\sim[2^{nR}]^L} \mathop{\mathbb{E}}_{(\mathbf{x}_{\mathbf{i}_1},\cdots,\mathbf{x}_{\mathbf{i}_L})\sim P_{\mathbf{x}_{\mathbf{i}_1},\cdots,\mathbf{x}_{\mathbf{i}_L}}} [Q_{\mathbf{i}_1,\cdots,\mathbf{i}_L}],$$

which is a contradiction and finishes the proof.

Upper bound on Q. By the assumption of list decodability, for every L-tuple of distinct codewords $\underline{x}_1, \cdots, \underline{x}_L \in \mathcal{C}'$, there is a codeword $\underline{x}_i$ ($i \in [L]$) among them such that

$$d_{\mathrm{H}}\left(\underline{x}_i, \mathsf{MAJ}\left(\underline{x}_1, \cdots, \underline{x}_L\right)\right) \geqslant np.$$

Equivalently,

$$\mathbb{E}\left[\mathbf{x}_i \mathsf{MAJ}\left(\mathbf{x}_1, \cdots, \mathbf{x}_L\right)\right] \leqslant \frac{r-\eta}{L}.$$

Since $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$ is symmetric and $\mathcal{C}'$ is $\left(\zeta, \widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}\right)$-equicoupled, we expect that the term $\mathbb{E}\left[\mathbf{x}_j \mathsf{MAJ}\left(\mathbf{x}_1, \cdots, \mathbf{x}_L\right)\right] \lesssim \frac{r-\eta}{L}$ for all $j \in [L]$, potentially with some slack depending on ζ. Indeed, for any $j \in [L] \backslash \{i\}$ (without loss of generality, assume $j > i$),

$$\begin{aligned}
&\left|\mathbb{E}\left[\mathbf{x}_i \mathsf{MAJ}\left(\mathbf{x}_1, \cdots, \mathbf{x}_L\right)\right] - \mathbb{E}\left[\mathbf{x}_j \mathsf{MAJ}\left(\mathbf{x}_1, \cdots, \mathbf{x}_L\right)\right]\right| \\
=&\left| \sum_{x_1,\cdots,x_L \in \{-1,1\}} \tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_1), \cdots, \phi^{-1}(x_L)) x_i \mathsf{MAJ}\left(x_1, \cdots, x_L\right) \right. \\
&\left. - \sum_{x_1,\cdots,x_L \in \{-1,1\}} \tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_1), \cdots, \phi^{-1}(x_L)) x_j \mathsf{MAJ}\left(x_1, \cdots, x_L\right) \right| \\
=&\left| \sum_{x_1,\cdots,x_L \in \{-1,1\}} \tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_1), \cdots, \phi^{-1}(x_L)) x_i \mathsf{MAJ}\left(x_1, \cdots, x_L\right) \right.
\end{aligned} \tag{112}$$

$$
\begin{aligned}
&- \sum_{x_{\sigma(1)},\cdots,x_{\sigma(L)}\in\{-1,1\}} \tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_{\sigma(1)}),\cdots,\phi^{-1}(x_{\sigma(L)}))x_{\sigma(j)}\mathsf{MAJ}\left(x_{\sigma(1)},\cdots,x_{\sigma(L)}\right)\Bigg| \qquad (113)\\
=&\Bigg|\sum_{x_1,\cdots,x_L\in\{-1,1\}} \tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_1),\cdots,\phi^{-1}(x_L))x_i\mathsf{MAJ}\left(x_1,\cdots,x_L\right)\\
&- \sum_{x_1,\cdots,x_L\in\{-1,1\}} \tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_{\sigma(1)}),\cdots,\phi^{-1}(x_{\sigma(L)}))x_i\mathsf{MAJ}\left(x_1,\cdots,x_L\right)\Bigg|\\
=&\left|\sum_{x_1,\cdots,x_L\in\{-1,1\}}\left[\begin{array}{c}\left(\begin{array}{c}\tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_1),\cdots,\phi^{-1}(x_L))\\ -\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(\phi^{-1}(x_1),\cdots,\phi^{-1}(x_L))\end{array}\right)\\ +\left(\begin{array}{c}\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(\phi^{-1}(x_{\sigma(1)}),\cdots,\phi^{-1}(x_{\sigma(L)}))\\ -\tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_{\sigma(1)}),\cdots,\phi^{-1}(x_\sigma(L)))\end{array}\right)\end{array}\right]x_i\mathsf{MAJ}\left(x_1,\cdots,x_L\right)\right| \qquad (114)\\
\leqslant&\left(\begin{array}{c}\left|\begin{array}{c}\tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_1),\cdots,\phi^{-1}(x_L))\\ -\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(\phi^{-1}(x_1),\cdots,\phi^{-1}(x_L))\end{array}\right|\\ +\left|\begin{array}{c}\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}(\phi^{-1}(x_{\sigma(1)}),\cdots,\phi^{-1}(x_{\sigma(L)}))\\ -\tau_{\underline{x}_1,\cdots,\underline{x}_L}(\phi^{-1}(x_{\sigma(1)}),\cdots,\phi^{-1}(x_\sigma(L)))\end{array}\right|\end{array}\right)\left|\sum_{x_1,\cdots,x_L\in\{-1,1\}} x_i\mathsf{MAJ}\left(x_1,\cdots,x_L\right)\right| \qquad (115)\\
\leqslant&2\zeta\cdot\frac{2^L}{L}\mathop{\mathbb{E}}_{(\mathbf{x}_1,\cdots,\mathbf{x}_L)\sim\{-1,1\}^L}\left[(\mathbf{x}_1+\cdots+\mathbf{x}_L)\,\mathsf{MAJ}\left(\mathbf{x}_1,\cdots,\mathbf{x}_L\right)\right] \qquad (116)\\
=&\frac{2^{L+1}}{L}\zeta\mathop{\mathbb{E}}_{(\mathbf{x}_1,\cdots,\mathbf{x}_L)\sim\{-1,1\}^L}\left[|\mathbf{x}_1+\cdots+\mathbf{x}_L|\right]\\
=&\frac{2^{L+1}r}{L}\zeta. \qquad (117)
\end{aligned}
$$

In the above chain of equalities and inequalities, we used the following facts.

1. In Eqn. (113), $\sigma\in S_L$ denotes the transposition which swaps the i-th and j-th element,
$$\sigma=\begin{pmatrix}1&\cdots&i-1&i&i+1&\cdots&j-1&j&j+1&\cdots&L\\1&\cdots&i-1&j&i+1&\cdots&j-1&i&j+1&\cdots&L\end{pmatrix}.$$
2. Eqn. (114) is due to symmetry of $\widehat{P}_{\mathbf{x}_1,\cdots,\mathbf{x}_L}$.
3. Inequality (115) is by triangle inequality of absolute value.
4. Eqn. (116) follows since
$$\left|\sum_{x_1,\cdots,x_L\in\{-1,1\}} x_i\mathsf{MAJ}\left(x_1,\cdots,x_L\right)\right|=2^L\left|\frac{1}{L}\sum_{i=1}^{L}\sum_{x_1,\cdots,x_L\in\{-1,1\}}\frac{1}{2^L}x_i\mathsf{MAJ}\left(x_1,\cdots,x_L\right)\right|,$$
and the expectation is over $\mathbf{x}_i$'s which are independent and uniformly distributed on $\{-1,1\}$.

Now, for any $j\in[L]\backslash\{i\}$,

$$
\begin{aligned}
\mathbb{E}\left[\mathbf{x}_j\mathsf{MAJ}\left(\mathbf{x}_1,\cdots,\mathbf{x}_L\right)\right]=&\mathbb{E}\left[\mathbf{x}_i\mathsf{MAJ}\left(\mathbf{x}_1,\cdots,\mathbf{x}_L\right)\right]+\left(\mathbb{E}\left[\mathbf{x}_j\mathsf{MAJ}\left(\mathbf{x}_1,\cdots,\mathbf{x}_L\right)\right]\right.\\
&\left.-\mathbb{E}\left[\mathbf{x}_i\mathsf{MAJ}\left(\mathbf{x}_1,\cdots,\mathbf{x}_L\right)\right]\right)\\
\leqslant&\frac{r-\eta}{L}+\frac{2^{L+1}r}{L}\zeta.
\end{aligned}
$$

Thus we have

$$\mathbb{E}\left[(\mathbf{x}_1+\cdots+\mathbf{x}_L)\,\mathsf{MAJ}\left(\mathbf{x}_1,\cdots,\mathbf{x}_L\right)\right]\leqslant r-\eta+\frac{2^{L+1}r(L-1)}{L}\zeta.$$

That is,

$$\begin{aligned}\mathbb{E}\left[Q_{1,\cdots,L}\right] =& \mathbb{E}\left[\left(\mathbf{x}_1+\cdots+\mathbf{x}_L\right)\mathsf{MAJ}\left(\mathbf{x}_1,\cdots,\mathbf{x}_L\right)-r\right]\\ \leqslant& -\eta+\frac{2^{L+1}r(L-1)}{L}\zeta\\ =& -\frac{\eta}{2},\end{aligned} \tag{118}$$

where the last Eqn. (118) follows by the choice of ζ (Eqn. (111)). Since the above calculations work for any list $\underline{x}_1,\cdots,\underline{x}_L\in\mathcal{C}'$ of *distinct* codewords, we have that for $(i_1,\cdots,i_L)\in\binom{[M']}{L}$, the same bound holds,

$$\mathbb{E}\left[Q_{i_1,\cdots,i_L}\right]\leqslant-\frac{\eta}{2}.$$

For lists $(i_1,\cdots,i_L)\in[M']^L$ that are not all distinct, we use the trivial bound,

$$\begin{aligned}\mathbb{E}\left[Q_{i_1,\cdots,i_L}\right] =& \mathbb{E}\left[\left|\mathbf{x}_{i_1}+\cdots+\mathbf{x}_{i_L}\right|-r\right]\\ \leqslant& L-r.\end{aligned}$$

Overall we have

$$\begin{aligned}Q=& \mathop{\mathbb{E}}_{(\mathbf{i}_1,\cdots,\mathbf{i}_L)\sim[2^{nR}]^L}\mathbb{E}\left[Q_{\mathbf{i}_1,\cdots,\mathbf{i}_L}\right]\\ =&\frac{1}{2^{nRL}}\left(\sum_{i_1,\cdots,i_L\in[2^{nR}]\text{ distinct}}Q_{i_1,\cdots,i_L}+\sum_{i_1,\cdots,i_L\in[2^{nR}]\text{ not distinct}}Q_{i_1,\cdots,i_L}\right)\\ \leqslant&\frac{1}{2^{nRL}}\Bigg[2^{nR}\left(2^{nR}-1\right)\cdots\left(2^{nR}-L+1\right)\left(-\frac{\eta}{2}\right)\\ &+\left(2^{nRL}-2^{nR}\left(2^{nR}-1\right)\cdots\left(2^{nR}-L+1\right)\right)(L-r)\Bigg]\\ <&0.\end{aligned} \tag{119}$$

The last inequality (119) holds if

$$\left|\mathcal{C}'\right|>\max\left\{2(L-1),\frac{2^{L+1}L!(L+r)}{\eta}\right\},$$

by similar calculations to Sec. 13.2.

Lower bound on Q. Following the calculations in the proof of generalized Plotkin bound for list decoding, we have

$$\begin{aligned}&Q+r\\ =&\mathop{\mathbb{E}}_{(\mathbf{i}_1,\cdots,\mathbf{i}_L)\sim[2^{nR}]^L}\mathbb{E}\left[\left|\mathbf{x}_{\mathbf{i}_1}+\cdots+\mathbf{x}_{\mathbf{i}_L}\right|\right]\\ =&\frac{1}{2^{nRL}}\sum_{i_1,\cdots,i_L\in[2^{nR}]}\sum_{x_1,\cdots,x_L\in\{-1,1\}}\tau_{\underline{x}_{i_1},\cdots,\underline{x}_{i_L}}\left(\phi^{-1}(x_1),\cdots,\phi^{-1}(x_L)\right)\left|x_1+\cdots+x_L\right|\\ =&\frac{1}{2^{nRL}}\sum_{\substack{i_1,\cdots,i_L\in\left[2^{nR}\right]\\ x_1,\cdots,x_L\in\{-1,1\}}}\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{\left\{\underline{x}_{i_1}(j)=\phi^{-1}(x_1)\right\}}\cdots\mathbb{1}_{\left\{\underline{x}_{i_L}(j)=\phi^{-1}(x_L)\right\}}\left|x_1+\cdots+x_L\right|\end{aligned} \tag{120}$$

$$= \frac{1}{n}\sum_{j=1}^{n}\sum_{x_1,\cdots,x_L\in\{-1,1\}}\prod_{\ell=1}^{L}\left(\frac{1}{2^{nR}}\sum_{i\in[2^{nR}]}\mathbb{1}_{\{\underline{x}_i(j)=\phi^{-1}(x_\ell)\}}\right)|x_1+\cdots+x_L| \tag{121}$$

$$= \frac{1}{n}\sum_{j=1}^{n}\sum_{x_1,\cdots,x_L\in\{-1,1\}}\prod_{\ell=1}^{L}P_{\mathbf{x}}^{(j)}(\phi^{-1}(x_\ell))\,|x_1+\cdots+x_L| \tag{122}$$

$$= \mathop{\mathbb{E}}_{\mathbf{j}\sim[n]}\left[\mathop{\mathbb{E}}_{\left(\mathbf{x}_1^{(\mathbf{j})},\cdots,\mathbf{x}_L^{(\mathbf{j})}\right)\sim\left(P_{\mathbf{x}}^{(\mathbf{j})}\right)^{\otimes L}}\left[\left|\mathbf{x}_1^{(\mathbf{j})}+\cdots+\mathbf{x}_L^{(\mathbf{j})}\right|\right]\right]. \tag{123}$$

In the above calculations, we used the following definitions and facts.

1. Eqn. (120) follows from the definition of joint types.
2. Eqn. (121) is obtained by rearranging terms.
3. In Eqn. (122), as before, we let, for $j\in[n]$, $x\in\mathbb{F}_2$,

$$P_{\mathbf{x}}^{(j)}(x)=\frac{1}{2^{nR}}\sum_{i\in[2^{nR}]}\mathbb{1}_{\{\underline{x}_i(j)=x\}}$$

denote the empirical distribution of the j-th *column* of $\mathcal{C}'$ when viewed as an $M'\times n$ matrix.

In expression (123), the j-th summand can be viewed as the translation distance of a non-lazy one-dimensional random walk after L steps. The walker moves left ($x=1$) with probability $P_{\mathbf{x}}^{(j)}(1)$ and moves right ($x=0$) with probability $P_{\mathbf{x}}^{(j)}(0)$. It is not hard to check that the expected translation distance is minimized when the walker is unbiased, i.e., when $P_{\mathbf{x}}^{(j)}(1)=P_{\mathbf{x}}^{(j)}(0)=1/2$. This is formally justified in Appendix C. Hence, for every $j\in[n]$,

$$\mathop{\mathbb{E}}_{\left(\mathbf{x}_1^{(j)},\cdots,\mathbf{x}_L^{(j)}\right)\sim\left(P_{\mathbf{x}}^{(j)}\right)^{\otimes L}}\left[\left|\mathbf{x}_1^{(j)}+\cdots+\mathbf{x}_L^{(j)}\right|\right]-r\geqslant 0.$$

Since the above bound is valid for every $j\in[n]$, it is still valid averaged over $\mathbf{j}\sim[n]$. Hence we have $Q\geqslant 0$. ◀

17 GV rate vs. cloud rate

In this section, we are concerned with the question of unique decoding (special case where $L-1=1$) under the bit-flip model.

In [43], bounds on achievable rates of codes for general adversarial channels are provided. A Gilbert–Varshamov-type expression was obtained using a purely random code construction, and a rate lower bound (we call *cloud rate*) that generalizes the GV-type expression was given by a cloud code construction. We evaluate both bounds under the bit-flip model. We show that the Gilbert–Varshamov-type bound for general adversarial channels indeed coincide with the classic GV bound in this particular setting. We also provide a convex program for evaluating the cloud rate.

We use the probability vector $\begin{bmatrix}P_{\mathbf{x}}(1) & \cdots & P_{\mathbf{x}}(|\mathcal{X}|)\end{bmatrix}^\top$ to denote a distribution $P_{\mathbf{x}}\in\Delta(\mathcal{X})$. Take any input distribution

$$P_{\mathbf{x}}=\mathrm{Bern}(w)=\begin{bmatrix}1-w\\ w\end{bmatrix},$$

from $\Delta(\{0,1\})$, we first explicitly compute the basic objects we are concerned with in this paper.

$$
\begin{aligned}
\Delta :=& \Delta(\{0,1\}) \\
=& \left\{ P_{\mathbf{x}_1,\mathbf{x}_2} \in \mathbb{R}^{2\times 2} \colon \begin{array}{rl} P_{\mathbf{x}_1,\mathbf{x}_2}(x_1,x_2) & \geqslant 0,\ \forall x_1,x_2 \\ \sum_{x_1,x_2} P_{\mathbf{x}_1,\mathbf{x}_2}(x_1,x_2) & = 1 \end{array} \right\} \\
=& \left\{ \begin{bmatrix} a & c \\ d & b \end{bmatrix} \in \mathbb{R}^{2\times 2} \colon \begin{array}{rl} a,b,c,d & \geqslant 0 \\ a+b+c+d & = 1 \end{array} \right\} \\
=& \left\{ \begin{bmatrix} a & c \\ 1-a-b-c & b \end{bmatrix} \in \mathbb{R}^{2\times 2} \colon \begin{array}{rl} a,b,c & \geqslant 0 \\ a+b+c & \leqslant 1 \end{array} \right\}.
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{J}(w) :=& \mathcal{J}\left(\begin{bmatrix} 1-w \\ w \end{bmatrix} \right) \\
=& \{ P_{\mathbf{x}_1,\mathbf{x}_2} \in \Delta \colon P_{\mathbf{x}_1} = P_{\mathbf{x}_2} = P_{\mathbf{x}} \} \\
=& \left\{ \begin{bmatrix} a & c \\ d & b \end{bmatrix} \in \mathbb{R}^{2\times 2} \colon \begin{array}{l} a,b,c,d \geqslant 0 \\ a+b+c+d = 1 \\ d+b = w \\ c+b = w \end{array} \right\} \\
=& \left\{ \begin{bmatrix} 1-w-d & d \\ d & w-d \end{bmatrix} \in \mathbb{R}^{2\times 2} \colon 0 \leqslant d \leqslant \min\{w, 1-w\} \right\}.
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{K}(w,p) :=& \mathcal{K}\left(\begin{bmatrix} 1-w \\ w \end{bmatrix} \right) \\
=& \{ P_{\mathbf{x}_1,\mathbf{x}_2} \in \mathcal{J}(w) : P_{\mathbf{x}_1,\mathbf{x}_2}(0,1) + P_{\mathbf{x}_1,\mathbf{x}_2}(1,0) \leqslant 2p \} \\
=& \left\{ \begin{bmatrix} 1-w-d & d \\ d & w-d \end{bmatrix} \in \mathbb{R}^{2\times 2} \colon 0 \leqslant d \leqslant \min\{w, 1-w, p\} \right\}.
\end{aligned}
$$

Since $\mathsf{CP}_2 = \mathsf{DNN}_2$, we have

$$
\begin{aligned}
&\mathsf{CP}_2(w) \\
=& \mathsf{CP}_2 \cap \mathcal{J}(w) \\
=& \left\{ \begin{bmatrix} w-d & d \\ d & 1-w-d \end{bmatrix} : 0 \leqslant d \leqslant \min\{w, 1-w\},\ (w-d)(1-w-d) - d^2 \geqslant 0 \right\} \\
=& \left\{ \begin{bmatrix} w-d & d \\ d & 1-w-d \end{bmatrix} : 0 \leqslant d \leqslant w - w^2 \right\}.
\end{aligned}
$$

Note that to ensure $\mathsf{CP}_2(w) \backslash \mathcal{K}(w,p) \neq \varnothing$, we need

$$
0 < p < 1/4,\ w \in \left(\tfrac{1-\sqrt{1-4p}}{2}, \tfrac{1+\sqrt{1-4p}}{2} \right).
$$

In other words, $0 < w < 1$ and $0 < p < w - w^2$. In this case,

$$
\mathcal{K}(w,p) = \left\{ \begin{bmatrix} 1-w-d & d \\ d & w-d \end{bmatrix} \in \mathbb{R}^{2\times 2} \colon 0 \leqslant d \leqslant p \right\}.
$$

Actually, if the above conditions hold, then when $1/3 \leqslant w < 1$, the boundary of $\mathcal{K}(w,p)$ is p and the boundary of $\mathsf{CP}_2(w)$ is $w - w^2$. Note that the right boundary $\begin{bmatrix} (1-w)^2 & w-w^2 \\ w-w^2 & w^2 \end{bmatrix} = \begin{bmatrix} 1-w \\ w \end{bmatrix}^{\otimes 2}$ of $\mathsf{CP}_2(w)$ is the *only* distribution in $\mathsf{CP}_2(w)$ of CP-rank-1.

GV rate. We first state the GV-type expression given by in [43].

▶ **Lemma 81** (Gilbert–Varshamov rate). *For a general adversarial channel given by* $\mathcal{A} = \{\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\}$*, its unique decoding capacity is at least*

$$R_{\mathrm{GV}} = \max_{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \min_{P_{\mathbf{x}_1,\mathbf{x}_2} \in \mathcal{K}(P_{\mathbf{x}})} I(\mathbf{x};\mathbf{x}'),$$

where the mutual information is calculated using $P_{\mathbf{x}_1,\mathbf{x}_2}$*.*

We now evaluate the above expression under the bit-flip model.

$$\begin{aligned}
R_{\mathrm{GV}} &= \max_{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \min_{P_{\mathbf{x}_1,\mathbf{x}_2} \in \mathcal{K}(P_{\mathbf{x}})} I(\mathbf{x};\mathbf{x}') \\
&= \max_{\begin{bmatrix} 1-w \\ w \end{bmatrix} \in \Delta} \min_{\begin{bmatrix} 1-w-d & d \\ d & w-d \end{bmatrix} \in \mathcal{K}(w,p)} D\left(\begin{bmatrix} 1-w-d & d \\ d & w-d \end{bmatrix} \middle\| \begin{bmatrix} 1-w \\ w \end{bmatrix}^{\otimes 2} \right) \\
&= \max_{0<w<1} \min_{0 \leqslant d \leqslant p} (w-d) \log \frac{w-d}{w^2} + 2d \log \frac{d}{w(1-w)} + (1-w-d) \log \frac{1-w-d}{(1-w)^2} \\
&= \max_{0<w<1} (w-p) \log \frac{w-p}{w^2} + 2p \log \frac{p}{w(1-w)} + (1-w-p) \log \frac{1-w-p}{(1-w)^2} \\
&= (1/2-p) \log \frac{1/2-p}{(1/2)^2} + 2p \log \frac{p}{(1/2)(1-1/2)} + (1-1/2-p) \log \frac{1-1/2-p}{(1-1/2)^2} \\
&= 1 - H(2p).
\end{aligned}$$

This matches the classic GV bound given a greedy volume packing argument.

Cloud rate. We now state the cloud rate expression given by [43].

▶ **Lemma 82** (Cloud rate).

For a general adversarial channel $\mathcal{A} = \{\mathcal{X}, \lambda_{\mathbf{x}}, \mathcal{S}, \lambda_{\mathbf{s}}, \mathcal{Y}, W_{\mathbf{y}|\mathbf{x},\mathbf{s}}\}$, its unique decoding capacity is at least

$$R_{\mathrm{cloud}} = \max_{P_{\mathbf{x}} \in \lambda_{\mathbf{x}}} \max_{P_{\mathbf{x}_1,\mathbf{x}_2} \in \mathsf{CP}_2(P_{\mathbf{x}}) \setminus \mathcal{K}(P_{\mathbf{x}})} \max_{\substack{P_{\mathbf{u}}, P_{\mathbf{x}|\mathbf{u}}: \\ \left[P_{\mathbf{u}} P_{\mathbf{x}|\mathbf{u}}^{\otimes 2}\right]_{\mathbf{x}_1,\mathbf{x}_2} = P_{\mathbf{x}_1,\mathbf{x}_2}}} \min_{P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_1} \in \mathcal{K}_{\mathrm{cloud}}(P_{\mathbf{u},\mathbf{x}})} D\left(P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2} \middle\| P_{\mathbf{u}} P_{\mathbf{x}|\mathbf{u}}^{\otimes 2}\right),$$

where

$$\mathcal{K}_{\mathrm{cloud}}(P_{\mathbf{u},\mathbf{x}}) := \left\{ [P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2,\mathbf{s}_1,\mathbf{s}_2,\mathbf{y}}]_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2} \in \Delta\left(\mathcal{U} \times \mathcal{X}^2\right) : \begin{array}{rl} P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2,\mathbf{s}_1,\mathbf{s}_2,\mathbf{y}} \in & \Delta\left(\mathcal{U} \times \mathcal{X}^2 \times \mathcal{S}^2 \times \mathcal{Y}\right) \\ P_{\mathbf{s}_1}, P_{\mathbf{s}_2} \in & \lambda_{\mathbf{s}} \\ P_{\mathbf{u},\mathbf{x}_1,\mathbf{s}_1,\mathbf{y}} = & P_{\mathbf{u},\mathbf{x}} P_{\mathbf{s}_1|\mathbf{u},\mathbf{x}_1} W_{\mathbf{y}|\mathbf{x}_1,\mathbf{s}_1} \\ P_{\mathbf{u},\mathbf{x}_2,\mathbf{s}_2,\mathbf{y}} = & P_{\mathbf{u},\mathbf{x}} P_{\mathbf{s}_2|\mathbf{u},\mathbf{x}_2} W_{\mathbf{y}|\mathbf{x}_2,\mathbf{s}_2} \end{array} \right\}.$$

▶ Remark 83. The reason that [43] has to define a different confusability set $\mathcal{K}_{\mathrm{cloud}}$ when cloud code is using is that as a part of the code design, the distributions $P_{\mathbf{u}}, P_{\mathbf{u}|\mathbf{x}}$ are revealed to every party, including the adversary, hence he may be able to inject noise patterns that are potentially more malicious compared with the case where he does not have such knowledge. We refer the readers to the proof in [43].

In the bit-flip setting, it is easy to verify that

$$\begin{aligned}\mathcal{K}_{\text{cloud}}(P_{\mathbf{u},\mathbf{x}}) &= \left\{ P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2} \in \Delta\left(\mathcal{U}\times\mathcal{X}^2\right) : \begin{array}{rl} P_{\mathbf{u},\mathbf{x}_1} = P_{\mathbf{u},\mathbf{x}_2} = & P_{\mathbf{u},\mathbf{x}} \\ P_{\mathbf{x}_1,\mathbf{x}_2}(0,1) + P_{\mathbf{x}_1,\mathbf{x}_2}(1,0) \leqslant & 2p \end{array} \right\} \\ &= \left\{ p \in \mathbb{R}^{|\mathcal{U}|\times 2\times 2} : \begin{array}{rl} p_{u,x_1,x_2} \geqslant & 0,\ \forall u,x_1,x_2 \\ \sum_{u,x_1,x_2} p_{u,x_1,x_2} = & 1 \\ \sum_{x_2} p_{u,x_1,x_2} = & p_{u,x_1},\ \forall u,x_1 \\ \sum_{x_1} p_{u,x_1,x_2} = & p_{u,x_2},\ \forall u,x_2 \\ \sum_u p_{u,0,1} + p_{u,1,0} \leqslant & 2p \end{array} \right\}.\end{aligned}$$

We use the notation $p_{u,x_1,x_2} \coloneqq P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}(u,x_1,x_2)$ and $p_{u,x} \coloneqq P_{\mathbf{u},\mathbf{x}}(u,x)$ for all $u \in \mathcal{U}, x_1, x_2 \in \{0,1\}$. The third maximization is over all extensions which correspond to CP decompositions of $P_{\mathbf{x}_1,\mathbf{x}_2}$. Note that for a CP matrix, its CP decomposition is not necessarily unique, even if we require the decomposition to meet the CP-rank [23]. A CP decomposition of a CP distribution can contain an arbitrarily large number of terms. Here we focus on decompositions which *meet* the CP-rank of $P_{\mathbf{x}_1,\mathbf{x}_2}$. That is, $|\mathcal{U}| = \text{CP-rank}(P_{\mathbf{x}_1,\mathbf{x}_2})$.

Note that the objective function KL-divergence also equals

$$D\left(P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2} \middle\| P_{\mathbf{u}} P_{\mathbf{x}|\mathbf{u}}^{\otimes 2}\right) = I\left(\mathbf{x}_1;\mathbf{x}_2|\mathbf{u}\right),$$

where the mutual information is w.r.t. $P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}$.

Note that even if we could show $R_{\text{cloud}} \leqslant R_{\text{GV}}$, this does *not* mean cloud codes will never attain a rate larger than the GV bound. It only means that the cloud rate expression we have cannot take values larger than the GV bound. This is because our bounds are only achievable, but we do not have matching upper bounds. Indeed, this is an extremely difficult question even under simple models.

Actually *all* CP decompositions meeting the CP-rank of a CP distribution can be computed. For a CP-rank-2 distribution $\begin{bmatrix} 1-w-b & b \\ b & w-b \end{bmatrix} \in \mathsf{CP}_2(w) \backslash \mathcal{K}(w,p)$ where $b \neq w - w^2$, we write its CP decomposition as

$$\begin{aligned}\begin{bmatrix} 1-w-b & b \\ b & w-b \end{bmatrix} =& \alpha \begin{bmatrix} 1-u \\ u \end{bmatrix}^{\otimes 2} + \beta \begin{bmatrix} 1-v \\ v \end{bmatrix}^{\otimes 2} \\ =& \begin{bmatrix} \alpha(1-u)^2 + \beta(1-v)^2 & \alpha u(1-u) + \beta v(1-v) \\ \alpha u(1-u) + \beta v(1-v) & \alpha u^2 + \beta v^2 \end{bmatrix}.\end{aligned}$$

Solving the equation in terms of b and u, we have

$$\begin{aligned}\alpha \coloneqq& \alpha(w,b,u) = \frac{w-b-w^2}{u^2+w-2uw-b}, \\ \beta \coloneqq& \beta(w,b,u) = 1-\alpha = \frac{(u-w)^2}{u^2+w-2uw-b}, \\ v \coloneqq& v(w,b,u) = \frac{b-w+uw}{w-u},\end{aligned}$$

where $u \in \left[0, \frac{b}{1-w}\right] \cup \left[\frac{w-b}{w}, 1\right]$.

Any such decomposition gives rise to a joint distribution $P_{\mathbf{u}} P_{\mathbf{x}|\mathbf{u}}^{\otimes 2}$ which is a $2\times 2\times 2$ tensor.

$$P_{\mathbf{u}=0} P_{\mathbf{x}|\mathbf{u}=0}^{\otimes 2} = \begin{bmatrix} \alpha(1-u)^2 & \alpha u(1-u) \\ \alpha u(1-u) & \alpha(1-u)^2 \end{bmatrix}, \quad P_{\mathbf{u}=1} P_{\mathbf{x}|\mathbf{u}=1}^{\otimes 2} = \begin{bmatrix} \beta v^2 & \beta v(1-v) \\ \beta v(1-v) & \beta(1-v)^2 \end{bmatrix}.$$

It also induces a distribution $P_{\mathbf{u},\mathbf{x}}$.

$$P_{\mathbf{u},\mathbf{x}} = \begin{bmatrix} \alpha(1-u) & \alpha u \\ \beta(1-v) & \beta v \end{bmatrix}.$$

Now for any CP decomposition $P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}$ of a CP distribution $P_{\mathbf{x}_1,\mathbf{x}_2} = \begin{bmatrix} w-b & b \\ b & 1-w-b \end{bmatrix}$, the inner minimization can be written as minimizing a convex function over a polytope.

$$\begin{array}{rl} \min_p & D(p\|P_{\mathbf{u}}P_{\mathbf{x}|\mathbf{u}}^{\otimes 2}) \\ \text{subject to} & p \in \mathcal{K}_{\text{cloud}}(P_{\mathbf{u},\mathbf{x}}) \end{array}.$$

It can be expanded in the following explicit form.

$$\begin{array}{rl} \min_p & p_{0,0,0}\log\frac{p_{0,0,0}}{\alpha(1-u)^2} + p_{0,0,1}\log\frac{p_{0,0,1}}{\alpha u(1-u)} + p_{0,1,0}\log\frac{p_{0,1,0}}{\alpha u(1-u)} + p_{0,1,1}\log\frac{p_{0,1,1}}{\alpha u^2} \\ & + p_{1,0,0}\log\frac{p_{1,0,0}}{\beta(1-v)^2} + p_{1,0,1}\log\frac{p_{1,0,1}}{\beta v(1-v)} + p_{1,1,0}\log\frac{p_{1,0,0}}{\beta v(1-v)} + p_{1,1,1}\log\frac{p_{1,1,1}}{\beta v^2} \\ \text{subject to} & \left.\begin{array}{l} p_{i,j,k} \geqslant 0,\ \forall i,j,k \\ \sum_{i,j,k} p_{i,j,k} = 1 \end{array}\right\} p \in \Delta\left(\{0,1\}^3\right) \\ & \left.\begin{array}{l} p_{0,0,0} + p_{0,0,1} = \alpha(1-u) \\ p_{0,1,0} + p_{0,1,1} = \alpha u \\ p_{1,0,0} + p_{1,0,1} = \beta(1-v) \\ p_{1,1,0} + p_{1,1,1} = \beta v \end{array}\right\} [P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}]_{\mathbf{u},\mathbf{x}_1} = P_{\mathbf{u},\mathbf{x}} \\ & \left.\begin{array}{l} p_{0,0,0} + p_{0,1,0} = \alpha(1-u) \\ p_{0,0,1} + p_{0,1,1} = \alpha u \\ p_{1,0,0} + p_{1,1,0} = \beta(1-v) \\ p_{1,0,1} + p_{1,1,1} = \beta v \end{array}\right\} [P_{\mathbf{u},\mathbf{x}_1,\mathbf{x}_2}]_{\mathbf{u},\mathbf{x}_2} = P_{\mathbf{u},\mathbf{x}} \\ & p_{0,0,1} + p_{0,1,0} + p_{1,0,1} + p_{1,1,0} \leqslant 2p. \end{array}$$

Note that it is implied by the given constraints that $p_{u,x_1,x_2} = p_{u,x_2,x_1}$. Also, the p.m.f. constraint $\sum_{u,x_1,x_2} p_{u,x_1,x_2} = 1$ is actually redundant. Hence the problem can be simplified as follows.

$$\begin{array}{rl} \min_{\underline{p}} & p_{0,0,0}\log\frac{p_{0,0,0}}{\alpha(1-u)^2} + 2p_{0,0,1}\log\frac{p_{0,0,1}}{\alpha u(1-u)} + p_{0,1,1}\log\frac{p_{0,1,1}}{\alpha u^2} \\ & + p_{1,0,0}\log\frac{p_{1,0,0}}{\beta(1-v)^2} + 2p_{1,0,1}\log\frac{p_{1,0,1}}{\beta v(1-v)} + p_{1,1,1}\log\frac{p_{1,1,1}}{\beta v^2} \\ \text{subject to} & -p_{i,j,k} \leqslant 0,\ \forall i,j,k \\ & p_{0,0,0} + p_{0,0,1} = \alpha u \\ & p_{0,0,1} + p_{0,1,1} = \alpha(1-u) \\ & p_{1,0,0} + p_{1,0,1} = \beta v \\ & p_{1,0,1} + p_{1,1,1} = \beta(1-v) \\ & p_{0,0,1} + p_{1,0,1} \leqslant p. \end{array}$$

Let $D^*(w,b,u)$ denote the optimal value of the above minimization. The final cloud rate is given by

$$\max_{0<w<1} \max_{p<b\leqslant w-w^2} \max_{u\in\left[0,\frac{b}{1-w}\right]\cup\left[\frac{w-b}{w},1\right]} D^*(w,b,u),$$

where the first maximization corresponds to finding the optimal input distribution $\begin{bmatrix} 1-w \\ w \end{bmatrix}$, second maximization corresponds to finding the optimal CP distribution $\begin{bmatrix} 1-w-b & b \\ b & w-b \end{bmatrix}$ outside $\mathcal{K}(w)$, and the third optimization corresponds to finding the optimal CP-decomposition $\alpha\begin{bmatrix} 1-u \\ u \end{bmatrix}^{\otimes 2} + \beta\begin{bmatrix} 1-v \\ v \end{bmatrix}^{\otimes 2}$ of the optimal CP distribution.

18 Concluding remarks and open problems

In this paper, we study the list decoding problem on general adversarial channels for both large and small list sizes. Given any channel, for large (yet constant) list sizes, we prove the list decoding theorem which identifies the fundamental limit of list decoding. For small (yet arbitrary universal constant) list sizes, we characterize when positive rate list decodable codes are possible.

Many open questions are left after this work is done. We list some of them for future study.

1. In this paper, we made no attempt towards understanding channels with arbitrary transition distributions $W_{\mathbf{y}|\mathbf{x},\mathbf{s}}$ (instead of only those corresponding to deterministic bivariate functions). Pushing our results to such a general setting remains an intriguing open question.
2. Other adversarial channels under further assumptions, e.g., online (causal) channels, channels with feedback, channels with bounded memory, etc., are less understood. There are results regarding each of these topics under very restricted models, e.g., bit-flips [13, 6], deletions [12], etc..
3. We do not have any nontrivial *upper* bound on $(L-1)$-list decoding capacity for general adversarial channels. Existing upper bounds for error correction codes seem tricky to generalize. A reasonable starting point might be to extend the classic Elias–Bassalygo bound [4] whose proof has a similar spirit as the Plotkin bound.
4. Given any adversarial channel, when we are "below the Plotkin point" (i.e., there are non-confusable CP distributions), can we construct *explicit* codes of positive rate? We know that random codes are list decodable w.h.p..

References

1 Rudolf Ahlswede. Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 25(3):239–252, 1973.

2 Noga Alon, Boris Bukh, and Yury Polyanskiy. List-decodable zero-rate codes. *IEEE Transactions on Information Theory*, 65(3):1657–1667, 2018.

3 Alexei Ashikhmin, Alexander Barg, and Simon Litsyn. A new upper bound on codes decodable into size-2 lists. In *Numbers, Information and Complexity*, pages 239–244. Springer, 2000.

4 L. A. Bassalygo. New upper boundes for error-correcting codes. *Problems of Information Transmisson*, 1:32–35, 1965.

5 Avraham Ben-Aroya, Dean Doron, and Amnon Ta-Shma. Near-Optimal Erasure List-Decodable Codes, 2018. URL: `https://eccc.weizmann.ac.il/report/2018/065/`.

6 Elwyn R Berlekamp. *Block coding with noiseless feedback*. PhD thesis, Massachusetts Institute of Technology, 1964.

7 Abhishek Bhowmick and Shachar Lovett. List decoding Reed-Muller codes over small fields. *arXiv preprint*, 2014. `arXiv:1407.3433`.

8 David Blackwell, Leo Breiman, and A. J. Thomasian. The Capacity of a Class of Channels. *Ann. of Mathematical Statistics*, 30(4):1229–1241, 1959.

9 Vladimir M Blinovsky. Bounds for codes in the case of list decoding of finite volume. *Problems of Information Transmission*, 22:7–19, 1986.

10 Vladimir M Blinovsky. Code bounds for multiple packings over a nonbinary finite alphabet. *Problems of Information Transmission*, 41:23–32, 2005.

11 Vladimir M Blinovsky. On the convexity of one coding-theory function. *Problems of Information Transmission*, 44:34–39, 2008.

12 Boris Bukh, Venkatesan Guruswami, and Johan Håstad. An improved bound on the fraction of correctable deletions. *IEEE Transactions on Information Theory*, 63(1):93–103, 2016.

13 Z. Chen, S. Jaggi, and M.Langberg. A Characterization of the Capacity of Online (causal) Binary Channels. In *Proc. ACM Symp. on Discrete Algorithms (SODA)*, Portland, U.S.A, June 2015.

14 Sean Clark. How to show $\sum_{i=1}^{\lceil L/2 \rceil} \frac{\binom{2i-2}{i-1}}{i} 2^{-2i} = 1/2 - 2^{-L-1}\binom{L}{(L-1)/2}$? Mathematics Stack Exchange, February 2019. URL: `https://math.stackexchange.com/q/3101258` (version: 2019-02-05).

15 G. D. Cohen, S. N. Litsyn, and G. Zemor. Upper bounds on generalized distances. *IEEE transactions on Information Theory*, 40:2090–2092, November 1994.

16 Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

17 Imre Csiszár and Prakash Narayan. The Capacity of the Arbitrarily Varying Channel Revisited : Positivity, Constraints. *IEEE transactions on Information Theory*, 34:181–193, 1988.

18 Philippe Delsarte. An algebraic approach to the association schemes of coding theory. Technical report, Philips Research Laboratories, 1973.

19 Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. List-decodable robust mean estimation and learning mixtures of spherical gaussians. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1047–1060. ACM, 2018.

20 Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. Nearly Optimal Pseudorandomness From Hardness. *ECCC preprint TR19-099*, 2019.

21 Peter Elias. List decoding for noisy channels. Technical report, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1957.

22 Edgar N Gilbert. A comparison of signalling alphabets. *The Bell system technical journal*, 31(3):504–522, 1952.

23 Patrick Groetzner and Mirjam Dür. A factorization method for completely positive matrices. *preprint*, 2018.

24 V. Guruswami. *List Decoding of Error Correcting Codes (Lecture Notes in Computer Science),*. Springer-Verlag, NY, 2004.

25 Venkatesan Guruswami. List decoding in average-case complexity and pseudorandomness. In *2006 IEEE Information Theory Workshop-ITW'06 Punta del Este*, pages 32–36. IEEE, 2006.

26 Venkatesan Guruswami and Srivatsan Narayanan. Combinatorial limitations of average-radius list decoding. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 591–606. Springer, 2013.

27 Christopher J Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.

28 Sushrut Karmalkar, Pravesh Kothari, and Adam Klivans. List-Decodable Linear Regression. *arXiv preprint*, 2019. `arXiv:1905.05679`.

29 A. Lapidoth and P. Narayan. Reliable Communication under Channel Uncertainty. *IEEE transactions on Information Theory*, 44:2148–2177, 1998.

30 Jessie MacWilliams. A theorem on the distribution of weights in a systematic code. *Bell System Technical Journal*, 42(1):79–94, 1963.

31 R. J. McEliece, E. R. Rodemich, H. Jr. Rumsey, and L. R. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE transactions on Information Theory*, 23, 1977.

32 Michael Navon and Alex Samorodnitsky. Linear programming bounds for codes via a covering argument. *Discrete & Computational Geometry*, 41(2):199, 2009.

33 Morris Plotkin. Binary codes with specified minimum distance. *IRE Transactions on Information Theory*, 6(4):445–450, 1960.

34 Yury Polyanskiy. Upper bound on list-decoding radius of binary codes. *IEEE Transactions on Information Theory*, 62(3):1119–1128, 2016.

35 Prasad Raghavendra and Morris Yau. List Decodable Learning via Sum of Squares. *arXiv preprint*, 2019. arXiv:1905.04660.

36 Atri Rudra and Mary Wootters. Every list-decodable code for high noise has abundant near-optimal rate puncturings. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, 2014.

37 Atri Rudra and Mary Wootters. It'll probably work out: improved list-decoding through random operations. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, 2015.

38 Atri Rudra and Mary Wootters. Average-radius list-recoverability of random linear codes. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2018.

39 Anand Sarwate. *Robust and adaptive communication under uncertain interference*. PhD thesis, University of California, Berkeley, 2008.

40 Michael Sipser and Daniel A Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996.

41 Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, 2017.

42 RR Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, SSSR*, 117:739–741, 1957.

43 Xishi (Nicholas) Wang, Amitalok J. Budkuley, Andrej Bogdanov, and Sidharth Jaggi. When are large codes possible for AVCs? In *IEEE International Symposium on Information Theory (ISIT), Paris*, pages 632–636. IEEE, 2019.

44 L. R. Welch, R. J. McEliece, and H. Jr. Rumsey. A low-rate improvement on the Elias bound. *IEEE transactions on Information Theory*, 23, 1974.

45 Mary Wootters. On the list decodability of random linear codes with large error rates. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, 2013.

46 John M Wozencraft. List decoding. *Quarterly Progress Report*, 48:90–95, 1958.

47 Victor Vasilievich Zyablov and Mark Semenovich Pinsker. List concatenated decoding. *Problemy Peredachi Informatsii*, 17(4):29–33, 1981.

A CP tensors and coP tensors

A.1 Tensor products

▶ **Definition 84** (Tensor product). *For two tensors* $A \in \mathsf{Ten}_n^{\otimes m}, B \in \mathsf{Ten}_n^{\otimes \ell}$, *Their* tensor product *is defined as*

$$A \otimes B := [A\,(i_1, \cdots, i_m)\,B\,(j_1, \cdots, j_\ell)] \in \mathsf{Ten}_n^{\otimes (m+\ell)}.$$

▶ **Definition 85** (Frobenius inner product, Frobenius norm). *For two tensors* $A, B \in \mathsf{Ten}_n^{\otimes m}$, *Their* inner product *is defined as*

$$\langle A, B\rangle := \sum_{i_1,\cdots,i_m \in [n]} A(i_1, \cdots, i_m) B(i_1, \cdots, i_m).$$

The Frobenius norm *is defined as* $\|A\|_F := \sqrt{\langle A, A\rangle}$.

▶ **Definition 86** (Hadamard product). *For two tensors* $A, B \in \mathsf{Ten}_n^{\otimes m}$, *Their* Hadamard product *is defined as*

$$A \circ B := [A(i_1, \cdots, i_m) B(i_1, \cdots, i_m)] \in \mathsf{Ten}_n^{\otimes m}.$$

A.2 Tensor decomposition

▶ **Definition 87** (Canonical decomposition). *For a tensor* $A \in \mathsf{Ten}_n^{\otimes m}$*, its* canonical decomposition *has form*

$$A = \sum_{j=1}^{r} \alpha_j \bigotimes_{i=1}^{m} \underline{x}_{j,i},$$

where each $\underline{x}_{j,i} \in \mathbb{S}_2^{n-1}$*. The smallest* r *for* A *to admit such a decomposition is called the* rank *of* A*. If* A *is symmetric, then*

$$A = \sum_{j=1}^{r} \alpha_j \underline{x}_j^{\otimes m}$$

is an analog of the eigendecomposition of symmetric matrices. The smallest r *is called the* symmetric rank *of* A*.*

▶ **Conjecture 88.** *For* $A \in \mathsf{Sym}_n^{\otimes m}$*,* $\mathrm{rank}(A) = \text{sym-rank}(A)$*.*

▶ Remark 89. It is known to be true if $\mathrm{rank}(A) \leqslant m$.

▶ **Definition 90** (Tucker decomposition). *For a tensor* $A \in \mathsf{Ten}_n^{\otimes m}$*, the* Tucker decomposition *has form*

$$A = \sum_{j_1=1}^{r_1} \cdots \sum_{j_m=1}^{r_m} \alpha_{j_1,\cdots,j_m} \bigotimes_{i=1}^{m} \underline{x}_{j_i,j}.$$

It is an analogy of the singular value decomposition of matrices.

A tensor $A \in \mathsf{Ten}_n^{\otimes m}$ has $n(m-1)^{n-1}$ eigenvalues. A may have non-real eigenvalues even if A is symmetric. If an eigenvector is real, then the corresponding eigenvalue is also real. Such eigenvalues are called *H-eigenvalues*. They always exist for even-order tensors.

A.3 Special tensors

▶ **Definition 91** (NN tensors). *A tensor is said to be* non-negative *if each of its entry is non-negative. The set of order-m dimension-n non-negative tensors is denoted by* $\mathsf{NN}_n^{\otimes m}$

▶ **Definition 92** (PSD tensors, PD Tensors). *For even* m*,* $A \in \mathsf{Ten}_n^{\otimes m}$ *is* positive semidefinite (PSD) *if* $\langle A, \underline{x}^{\otimes m} \rangle \geqslant 0$ *for any* $\underline{x} \in \mathbb{R}^n$*.* A *is* positive definite (PD) *if the above inequality is strict for all* $\underline{x} \neq 0$*.*

The sets of PSD *and* PD *tensors is denoted by* $\mathsf{PSD}_n^{\otimes m}$ *and* $\mathsf{PD}_n^{\otimes m}$*, respectively.*

▶ **Definition 93** (CP tensors, CP tensor rank). *A tensor* $P \in \mathsf{Ten}_n^{\otimes m}$ *is said to be* completely positive *if for some* $r \geqslant 1$*, there are component-wise non-negative vectors* $\underline{p}_1, \cdots, \underline{p}_r \in \mathbb{R}^n_{\geqslant 0}$ *such that*

$$P = \sum_{j=1}^{r} \underline{p}_j^{\otimes m}.$$

The set of CP *tensors is denoted by* $\mathsf{CP}_n^{\otimes m}$*. The least* r *such that* P *has a completely positive decomposition is called the* CP-rank *of* P*. If* $\mathrm{span}\{P_1, \cdots, P_r\} = \mathbb{R}^n$ *then* P *is said to be* strongly CP*.*

▶ **Fact 94.** *Verifying if a symmetric non-negative tensor is* CP *is* NP*-hard.*

▶ **Definition 95** (coP tensors). *$A \in \mathsf{Sym}_n^{\otimes m}$ is* copositive *if $\langle A, \underline{x}^{\otimes}\rangle \geqslant 0$ for all $\underline{x} \in \mathbb{R}^n_{\geqslant 0}$. The set of copositive tensors is denoted by $\mathsf{coP}_n^{\otimes m}$.*

▶ **Theorem 96** (Duality). *$\mathsf{CP}_n^{\otimes m}$ and $\mathsf{coP}_n^{\otimes m}$ are closed convex pointed cones with nonempty interior in $\mathsf{Sym}_n^{\otimes m}$. For $m \geqslant 2$, $n \geqslant 1$, they are dual to each other.*

▶ **Definition 97** (DNN tensors). *For even m, $A \in \mathsf{Sym}_n^{\otimes m}$ is* doubly non-negative (DNN) *if A is entry-wise non-negative and $\langle A, \underline{x}^{\otimes m}\rangle$ is a sum-of-square as a polynomial in the components of $\underline{x}$.*

▶ **Fact 98.** *The double non-negativity of a tensor can be verified in polynomial time using SDP.*

▶ **Fact 99.** *The following inclusion relations between different sets of special tensors hold.*
1. $\mathsf{PSD}_n^{\otimes m} \subseteq \mathsf{coP}_n^{\otimes m}$.
2. $\mathsf{CP}_n^{\otimes m} \subseteq \mathsf{DNN}_n^{\otimes m} \subseteq \mathsf{NN}_n^{\otimes m} \subseteq \mathsf{coP}_n^{\otimes m} \subseteq \mathsf{Sym}_n^{\otimes m}$.

B Hypergraph Ramsey numbers

Let $R_k^{(r)}(s_1, \cdots, s_k)$ denote the smallest size of an r-uniform hypergraph such that for any k-colouring, there must be a monochromatic clique of size s_i for some $i \in [k]$.

Define tower function $t_1(x) = x$ and $t_{i+1}(x) = 2^{t_i(x)}$.

▶ **Lemma 100** (Properties of hypergraph Ramsey numbers). 1. *For any $i \in [k]$, and $s_j \geqslant r$ $(j \neq i)$,*

$$R_k^{(r)}(s_1, \cdots, s_{i-1}, r, s_{i+1}, \cdots, s_k) = R_{k-1}^{(r)}(s_1, \cdots, s_{i-1}, s_{i+1}, \cdots, s_k).$$

2. *For any $\sigma \in S_k$,*

$$R_k^{(r)}(s_1, \cdots, s_k) = R_k^{(r)}(s_{\sigma(1)}, \cdots, s_{\sigma(k)}).$$

▶ **Lemma 101** (Finiteness of hypergraph Ramsey numbers). *For any positive integers r, k, s_1, $\cdots$, s_k, the hypergraph Ramsey number $R_k^{(r)}(s_1, \cdots, s_k)$ is finite. In particular, it satisfies the following recursive inequalities.*

$$R_k^{(r)}(s_1, \cdots, s_k) \leqslant 1 + R_k^{(r-1)}\Big(R_k^{(r)}(s_1 - 1, s_2, \cdots, s_k), R_k^{(r)}(s_1, s_2 - 1, \cdots, s_k), \cdots, \cdots, R_k^{(r)}(s_1, s_2, \cdots, s_k - 1)\Big),$$

$$R_k^{(r)}(s_1, \cdots, s_k) \leqslant 1 + \sum_{i=1}^{k} R_k^{(r-1)}\Big(R_k^{(r)}(s_1, \cdots, s_{i-1}, s_i - 1, s_{i+1}, \cdots, s_k), \cdots, \cdots, R_k^{(r)}(s_1, \cdots, s_{i-1}, s_i - 1, s_{i+1}, \cdots, s_k)\Big),$$

$$R_k^{(r)}(s_1, \cdots, s_k) \leqslant R_{k-1}^{(r)}\left(s_1, \cdots, s_{k-2}, R_2^{(r)}(s_{k-1}, s_k)\right),$$

▶ **Lemma 102** (Bounds on hypergraph Ramsey numbers)**.**

1. *For any* s, t*,*

$$R_2^{(r)}(s,t) \leqslant 2^{\binom{R_2^{(r-1)}(s-1,t-1)}{r-1}}.$$

2. *For* $r \geqslant 3$*, there are constants* $c, c' > 0$ *such that*

$$t_{r-1}(c \cdot s^2) \leqslant R_2^{(r)}(s,s) \leqslant t_r(c' \cdot s).$$

3. *For* $s > k \geqslant 2$*, there are constants* $c, c' > 0$ *such that*

$$t_r(c \cdot k) < R_k^{(r)}(s, \cdots, s) < t_r(c' \cdot k \log k).$$

C Expected translation distance of a one-dimensional random walk

▶ **Lemma 103.** *Consider a random walk* $\mathbf{x}_1, \cdots, \mathbf{x}_L$ *of length* L*. Each* $\mathbf{x}_i$ $(1 \leqslant i \leqslant L)$ *is an independent and identically distributed* $\{-1, 1\}$*-valued random variable satisfying*

$$\Pr[\mathbf{x}_i = 1] = p, \quad \Pr[\mathbf{x}_i = -1] = 1 - p.$$

Without loss of generality, assume $p \geqslant 1/2$*. Then, we have that the expected translation distance* $\mathbb{E}[|\mathbf{x}_1 + \cdots + \mathbf{x}_L|]$ *of this random walk after* L *steps is minimized when* $p = 1/2$*.*

Proof. Create another walk $\mathbf{x}'_1, \cdots, \mathbf{x}'_L$ with $p = 1/2$ that is coupled with $\mathbf{x}_1, \cdots, \mathbf{x}_L$ in the following way.

$$\Pr\left[\mathbf{x}_i = 1 | \mathbf{x}'_i = 1\right] = 1, \quad \Pr\left[\mathbf{x}_i = 1 | \mathbf{x}'_i = -1\right] = 2p - 1.$$

It is easy to see that the distribution of $\mathbf{x}_1, \cdots, \mathbf{x}_L$ is preserved under this coupling.

$$\begin{aligned}\Pr[\mathbf{x}_i = 1] =& \Pr\left[\mathbf{x}'_i = 1\right] \Pr\left[\mathbf{x}_i = 1 | \mathbf{x}'_i = 1\right] + \Pr\left[\mathbf{x}'_i = -1\right] \Pr\left[\mathbf{x}_i = 1 | \mathbf{x}'_i = -1\right] \\ =& \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (2p - 1) \\ =& p.\end{aligned}$$

Now,

$$\begin{aligned}&\mathbb{E}[|\mathbf{x}_1 + \cdots + \mathbf{x}_L|] - \mathbb{E}\left[|\mathbf{x}'_1 + \cdots + \mathbf{x}'_L|\right] \\ =& \sum_{d \in \{-L, -L+2, \cdots, L-2, L\}} \\ & \qquad \sum_{\substack{x_1, \cdots, x_L \in \{-1,1\} \\ \sum_{i=1}^L x_i = d}} \Pr\left[\mathbf{x}'_1 = x_1, \cdots, \mathbf{x}'_L = x_L\right] \mathbb{E}\left[\left|\sum_{i=1}^L \mathbf{x}_i\right| - |d| \,\middle|\, \mathbf{x}'_1 = x_1, \cdots, \mathbf{x}'_L = x_L\right].\end{aligned}$$

For each translation distance $d \in \{-L, -L+2, \cdots, L-2, L\}$ and trajectory $x_1, \cdots, x_L \in \{-1, 1\}$ such that $\sum_{i=1}^L x_i = d$, let $\ell := \{i \in [L] \colon x_i = -1\}$. Note $2(d + \ell) = L$. We have

$$\begin{aligned}\mathbb{E}\left[\left|\sum_{i=1}^L \mathbf{x}_i\right| - |d| \,\middle|\, \mathbf{x}'_1 = x_1, \cdots, \mathbf{x}'_L = x_L\right] =& ((2p-1) \cdot 1 + (1 - (2p-1)) \cdot (-1))\ell - (-\ell) \\ =& 2(2p-1)\ell,\end{aligned}$$

which is non-negative and attains its minima 0 when $p = 1/2$. This finishes the proof. ◀

D Blinovsky [9] vs. Alon–Bukh–Polyanskiy [2]

In this section we show that, though differing ostensibly, the formulas of the Plotkin points for $(p, L-1)$-list decoding given by Blinovsky and Alon–Bukh–Polyanskiy actually agree with each other. The proof is essentially due to the user `Sean Clark` on `Mathematics Stack Exchange` [14].

For $L = 2k$ or $2k+1$ for some positive integer $k \in \mathbb{Z}_{>0}$, Blinovsky's formula is

$$P_{L-1} = \sum_{i=1}^{k} \frac{\binom{2(i-1)}{i-1}}{i} 2^{-2i};$$

while Alon–Bukh–Polyanskiy wrote it as

$$P_{L-1} = \frac{1}{2} - 2^{-2k-1}\binom{2k}{k}.$$

We are going to show that

▶ **Lemma 104.** *For any $k \geqslant 1$,*

$$\sum_{i=1}^{k} \frac{\binom{2(i-1)}{i-1}}{i} 2^{-2i} = \frac{1}{2} - 2^{-2k-1}\binom{2k}{k}.$$

Proof. To see the above two expressions are always evaluated to the same value, we first massage the above equation. Multiplying 2^{2k+2} on both sides, shifting the summation index and rearranging terms, we have

$$\sum_{i=0}^{k-1} \frac{\binom{2i}{i}}{i+1} 2^{2(k-i)} = 2^{2k+1} - 2\binom{2k}{k}.$$

Adding $\frac{\binom{2k}{k}}{k+1}$ on both sides, we get

$$\begin{aligned}
\sum_{i=0}^{k} \frac{\binom{2i}{i}}{i+1} 2^{2(k-i)} =& 2^{2k+1} - \left(2 - \frac{1}{k+1}\right)\binom{2k}{k} \\
=& 2^{2k+1} - \frac{2k+1}{k+1}\binom{2k}{k} \\
=& 2^{2k+1} - \binom{2k+1}{k+1} && (124) \\
=& 2^{2k+1} - \binom{2k+1}{k}, && (125)
\end{aligned}$$

where Eqn. (124) is by Fact (16) and Eqn. (125) is by Fact (15).

To show

$$\sum_{i=0}^{k} \frac{\binom{2i}{i}}{i+1} 2^{2(k-i)} = 2^{2k+1} - \binom{2k+1}{k}, \tag{126}$$

we conduct induction on k.

1. When $k = 0$, LHS $= 1 =$ RHS.
2. Assume (126) holds for certain $k \geqslant 1$. We want to show it also holds for $k+1$.

$$\begin{aligned}
\sum_{i=0}^{k+1} \frac{\binom{2i}{i}}{i+1} 2^{2(k+1-i)} =& 2^2 \sum_{i=0}^{k} \frac{\binom{2i}{i}}{i+1} 2^{2(k-i)} + \frac{\binom{2(k+1)}{k+1}}{k+2} \\
=& 2^2 \left(2^{2k+1} - \binom{2k+1}{k} \right) + \frac{\binom{2k+2}{k+1}}{k+2} && (127) \\
=& 2^{2(k+1)+1} - 2 \left(\binom{2k+1}{k} + \binom{2k+1}{k+1} \right) + \frac{\binom{2k+2}{k+1}}{k+2} && (128) \\
=& 2^{2(k+1)+1} - \left(2 - \frac{1}{k+2} \right) \binom{2k+2}{k+1} && (129) \\
=& 2^{2(k+1)+1} - \frac{2k+3}{k+2} \binom{2k+2}{k+1} \\
=& 2^{2(k+1)+1} - \binom{2(k+1)+1}{(k+1)+1}. && (130)
\end{aligned}$$

Eqn. (127), (128), (129) and (130) follow from induction hypothesis, Fact (15), Fact (17) and Fact (16), respectively. Hence Eqn. (126) holds for $k+1$ as well. ◀