List Decoding of Tanner and Expander Amplified Codes from Distance Certificates

Fernando Granha Jeronimo[*]   Shashank Srivastava[†]   Madhur Tulsiani[‡]

We develop new list decoding algorithms for Tanner codes and distance-amplified codes based on bipartite spectral expanders. We show that proofs exhibiting lower bounds on the minimum distance of these codes can be used as certificates discoverable by relaxations in the Sum-of-Squares (SoS) semidefinite programming hierarchy. Combining these certificates with certain entropic proxies to ensure that the solutions to the relaxations cover the entire list, then leads to algorithms for list decoding several families of codes up to the Johnson bound. We prove the following results:

- We show that the LDPC Tanner codes of Sipser-Spielman [IEEE Trans. Inf. Theory 1996] and Zémor [IEEE Trans. Inf. Theory 2001] with alphabet size q, block-length n and distance δ, based on an expander graph with degree d, can be list-decoded up to distance $\mathcal{J}_q(\delta) - \varepsilon$ in time $n^{O_{d,q}(1/\varepsilon^4)}$, where $\mathcal{J}_q(\delta)$ denotes the Johnson bound.

- We show that the codes obtained via the expander-based distance amplification procedure of Alon, Edmonds and Luby [FOCS 1995] can be list-decoded close to the Johnson bound using the SoS hierarchy, by reducing the list decoding problem to unique decoding of the base code. In particular, starting from *any* base code unique-decodable up to distance δ, one can obtain near-MDS codes with rate R and distance $1 - R - \varepsilon$, list-decodable up to the Johnson bound in time $n^{O_{\varepsilon,\delta}(1)}$.

- We show that the locally testable codes of Dinur et al. [STOC 2022] with alphabet size q, block-length n and distance δ based on a square Cayley complex with generator sets of size d, can be list-decoded up to distance $\mathcal{J}_q(\delta) - \varepsilon$ in time $n^{O_{d,q}(1/\varepsilon^4)}$, where $\mathcal{J}_q(\delta)$ denotes the Johnson bound.

Contents

1 Introduction

Expander graphs have been a powerful tool for the construction of codes with several interesting properties, and a variety of applications. A (very) small subsample of the list of applications already includes the seminal constructions of expander codes [SS96, Zém01], widely used distance amplification constructions [ABN$^+$92, AEL95], as well as recent breakthrough constructions of ε-balanced codes [TS17], locally testable codes [DEL$^+$22], and quantum LDPC codes [PK22, LZ22]. A detailed account of the rich interactions between coding theory and expander graphs, and pseudorandom objects in general, can be found in several excellent surveys and textbooks on these areas [Gur04, Vad12, HLW06, GRS23].

The combinatorial and spectral structure of codes based on expander graphs often leads to very efficient algorithms for unique-decoding. However, obtaining list-decoding for constructions based on expanders often require incorporating additional algebraic structure in the construction, to take advantage of the significant machinery for list-decoding using polynomials [Gur07]. While there are certainly important counterexamples to the above statement, such as the expander-based codes of Guruswami and Indyk [GI03] and Ta-Shma [TS17] which allow for list-decoding, we know of few *general techniques to exploit expansion for list-decoding.* In this work, we consider the question of finding techniques for list-decoding from errors, which can work in settings where no algebraic structure may be available, such as the decoding of LDPC codes constructed from expander graphs.

Building on the significant body of work for LP decoding of expander codes [Fel03, FWK05], we consider the question of decoding as an optimization problem, which can be approached via convex relaxations. We show that stronger relaxations obtained via the Sum-of-Squares (SoS) hierarchy of semidefinite programs, can in fact be used to obtain list-decoding algorithms for several code constructions based on expanders. These hierarchies can be viewed as proof systems [FKP19], with relaxations at a level t of the hierarchy corresponding to proofs which can be carried out by reasoning about sum-of-squares of polynomials of degree at most t in the optimization variables. The proof system corresponding to a small number of levels of the SoS hierarchy turns out to be powerful enough to capture the distance proofs for several expander-based codes, when the proofs rely on spectral properties of expander graphs. Combined with generic "covering lemmas" which ensure that the solutions to these relaxations include sufficient information about all codewords in the list, these can be used to design list-decoding algorithms for several families of codes based on expanders, up to the Johnson bound where the list size is known to be bounded.

1.1 Tanner Codes

Low-Density Parity Check (LDPC) codes were introduced by a foundational work of Gallager [Gal62] and graph-based constructions with lower bounds on the distance (based on girth), were obtained by Tanner [Tan81]. Sipser and Spielman [SS96] gave the first constructions of Tanner codes with distance bounds based on the expansion of the graph, which also admitted linear time encoding and (unique) decoding algorithms. An elegant construction based on bipartite spectral expanders, with particularly simple (linear-time) unique-decoding algorithms, was given by Zémor [Zém01]. Variants of these constructions have led to applications [RU08] and have also been used as building blocks in the recent constructions of locally testable codes by Dinur et al. [DEL$^+$22] and quantum LDPC codes by Panteleev and Kalachev [PK22] (see also [LZ22]).

There exist highly efficient algorithms for the unique-decoding of these codes from both probabilistic and adversarial errors, based on combinatorial arguments, linear programming relaxations [Fel03, ADS12, FWK05] and message passing algorithms [Gur06, RU08]. In the setting of *erasures* where the location of the corruptions in the transmitted codeword is known, recent work has also led to linear-time list decoding algorithms [RZWZ21, HW18], which also work for the more general task of list recovery in the large alphabet (high-rate) case [HW18]. However, to the best of our knowledge, no list decoding algorithms are known in the more challenging (and common) setting of *errors* when the location of the corruptions are unknown, even though random ensembles of LDPC codes are even known to combinatorially achieve list-decoding capacity [MRRZ$^+$20], and

thus have bounded list sizes up to optimal error radii.

We show that relaxations obtained via the SoS hierarchy can be used list-decode Zémor's construction of Tanner codes [Zém01], up to the Johnson bound (which is an error-radius where list sizes are always known to be bounded). Our proof technique can also be extended to work for other constructions of Tanner codes where the proof for the distance of the code is based on spectral arguments, but is easiest to illustrate in the context of Zémor's construction. We briefly recall the construction before describing our result.

Given a bipartite d-regular graph $G = (L, R, E)$ with $|L| = |R| = n$, Zémor's construction yields a code with block-length $m = |E| = nd$. Given an alphabet size q, the code consists of all edge-labelings $f \in [q]^m$, such that the labels in the neighborhood of every vertex [1], belong to a "base code" $\mathcal{C}_0 \subseteq [q]^d$. When the base code has (fractional) distance δ_0 and G has (normalized) second singular value at most λ for the biadjacency matrix, the distance of the code is known to be at least $\delta = \delta_0 \cdot (\delta_0 - \lambda)$. The Johnson bound for distance δ and alphabet size q is defined as $\mathcal{J}_q(\delta) := (1 - 1/q) \cdot \left(1 - (1 - q \cdot \delta/(q-1))^{1/2}\right)$, and is always greater than the unique-decoding radius $\delta/2$. We prove the following.

Theorem 1.1. *[Informal version of Theorem 7.1] Given a Tanner code $\mathcal{C}$ as above and $\varepsilon > 0$, there is a deterministic algorithm based on $q^{O(d)}/\varepsilon^4$ levels of the SoS hierarchy, which given an arbitrary $g \in [q]^m$, runs in time $n^{q^{O(d)}/\varepsilon^4}$, and recovers the list of codewords within distance $\mathcal{J}_q(\delta) - \varepsilon$ of g.*

Note that one can think of q, d in Zémor's construction and ε above as constants (in fact d is required to be constant for LDPC codes), in which case the above running time is polynomial in n. Of course, these running times are no match for the linear-time unique decoding, and erasure list-decoding, algorithms available for these codes. However, we view the techniques used in the proof of the above algorithm as a first step towards identifying the right structures, and designing truly efficient algorithms, to take advantage of expansion for list-decoding of LDPC codes (as has proved to be the case for several SoS-based algorithms in the past).

Our techniques also extend to yield a similar statement for the recent construction of locally testable codes by Dinur et al. [DEL+22], which are Tanner codes on a different structure called a "square Cayley complex". These are constructed using a group H and two generator sets $A, B \subseteq H$, with each generator set individually defining an expanding Cayley graph on H (with second eigenvalue bounded by λ). The sizes of the generator sets (equal to the graph degree) are taken as constant, say $|A| = |B| = d$. The construction relies on base codes $\mathcal{C}_A, \mathcal{C}_B \subseteq [q]^d$, with distances, say δ_A and δ_B, and is known to have distance at least $\delta = \delta_A \cdot \delta_B \cdot (\max\{\delta_A, \delta_B\} - \lambda)$.

Theorem 1.2. *[Informal version of Theorem 9.1] Given a code $\mathcal{C}^{SCC}$ with block length m and alphabet $[q]$, supported on a square Cayley complex as described above, and $\varepsilon > 0$, there is a deterministic algorithm based on $q^{O(d^2)}/\varepsilon^{O(1)}$ levels of the SoS hierarchy, which given an arbitrary $g \in [q]^m$, recovers the list of codewords within distance $\mathcal{J}_q(\delta) - \varepsilon$ of g.*

1.2 Distance Amplified Codes

The proofs for the distance of the above Tanner codes, are also very similar to the ones used for analyzing the distance amplification procedure of Alon, Edmonds, and Luby [AEL95] (AEL), based on expander graphs. While there several variants of this construction discussed in the literature, we will discuss a version of the AEL construction [Kop16] which is particularly close to the Tanner code construction of Zémor. Given a d-regular bipartite graph $G = (L, R, E)$ with second singular value λ, an "outer" code $\mathcal{C}_1 \subseteq [q_1]^n$ with distance δ_1, and an "inner" code $\mathcal{C}_0 \subseteq [q]^d$ with $|\mathcal{C}_0| = q_1$, the AEL procedure constructs a new code $\mathcal{C}^{AEL} \subseteq [q^d]^n$ with distance at least $\delta_0 - \frac{\lambda}{\delta_1}$. Thus, it yields constructions with arbitrarily large block lengths that inherit the parameters of the small inner code.

---

[1]One can also consider variants where the base code $\mathcal{C}_0$ is different for different vertices, but this does not make a difference for our purposes.

The AEL procedure has been used as an important ingredient for obtaining optimal rate-distance tradeoffs in several constructions, such as the capacity-achieving list-decodable codes by Guruswami and Rudra [GR06]. The amplification is achieved via a simple redistribution of symbols using the expander, and the construction also preserves several interesting local properties of the outer code, such as the property of being LDPC, or locally testable, or locally correctable [KMRZS16, GKO$^+$17]. We refer the reader to the discussion in [KMRZS16] for an excellent account of the applications and properties of the AEL construction.

The AEL procedure has been used to construct several list-decodable codes, including some of the results cited above, and a quantum analogue of the construction was also used recently by Bergamaschi et al. [BGG22] to obtain quantum codes meeting the Singleton bound (via quantum list decoding). However, for the resulting code $\mathcal{C}^{AEL}$ to be list-decodable, one often needs to assume stronger properties such as list-recovery, or some algebraic structure (or both) for the outer code $\mathcal{C}_1$. Since these stronger properties may not always be available (for example, when one wants to preserve some local properties for C_1 like being LDPC), we again consider the question of finding techniques which can allow for list-decoding $\mathcal{C}^{AEL}$ for expanding graphs G, without relying on additional structure from $\mathcal{C}_1$.

We show that relaxations based on the SoS hierarchy, can be used to list-decode the distance-amplified code $\mathcal{C}^{AEL}$, even when the outer code is only assumed to be *unique-decodable*. In particular, we prove the following result:

Theorem 1.3. *[Informal version of Theorem 7.2] Let $\mathcal{C}^{AEL}$ be a distance-amplified code as above, with the outer code $\mathcal{C}_1$ taken to be unique-decodable from radius δ_{dec} in time $T(n)$, and let $\delta = \delta_0 - \frac{\lambda}{\delta_{dec}}$. Then, for every $\varepsilon > 0$, there is a deterministic algorithm based on $q^{O(d)}/\varepsilon^4$ levels of the SoS hierarchy, which given an arbitrary $g \in [q^d]^n$, runs in time $n^{q^{O(d)}/\varepsilon^4} + O(T(n))$, and recovers the list of codewords within distance $\mathcal{J}_{q^d}(\delta) - \varepsilon$ of g.*

We note that the decoding radius for the above algorithm is $\mathcal{J}\left(\delta_0 - \frac{\lambda}{\delta_{dec}}\right)$ instead of $\mathcal{J}\left(\delta_0 - \frac{\lambda}{\delta_1}\right)$, which would be the Johnson bound for the true distance of the code $\mathcal{C}^{AEL}$. However, in applications of AEL, one often chooses parameters so that the distance of code is about δ_0, and the effect of the second term is minimized by choosing a small λ. When $\mathcal{C}_1$ is known to unique-decodable up to a smaller radius δ_{dec}, one can still obtain list-decodable codes up to (nearly) the Johnson bound by choosing G to be a sufficiently good expander (with small λ).

1.3   Our techniques

As mentioned earlier, our techniques are based on using the Sum-of-Squares hierarchy of convex relaxations for an optimization problem related to the decoding problem. In unique-decoding algorithms based on the LP relaxations, the optimization objective is to find the closest codeword to a given received word, and the correctness of the decoding procedure often relies on the LP being integral for an appropriate range of parameters. In contrast, algorithms for list-decoding actually need to ensure that the solution to the convex relaxation has sufficient information about all codewords in the list, and so it is important that the solution is *not* integral but rather a "maximally-convex" combination, covering all of the list elements. This can be ensured by statements which we call "covering lemmas", which are discussed in more detail in Section 4. The proofs for the covering lemmas are based on the techniques from [AJQ$^+$20], where these were used for the list-decoding of direct-sum codes.

A second key component of our proof, which makes the SoS hierarchy particularly appealing to work with, is for the relaxation to be able to capture global properties of the code, such as the distance. While local properties of the code, such as the structures of the base/inner codes are enforced through explicit constraints included in the relaxation, the global property of distance is a nontrivial consequence of these constraints. However, the *proofs* of these distance properties are spectral in nature, for the codes we consider here, which makes them discoverable by the SoS hierarchy.

For example, the proofs rely on statements such as the expander mixing lemma, which can viewed as a consequence of statements like $\langle f, Af \rangle \leq \lambda \cdot \|f\|^2$ when $\langle f, 1 \rangle = 0$, and A is the (normalized) adjacency matrix of a graph with second eigenvalue at most λ. Taking Π to be the projector to the space orthogonal to the all-ones vector, we can re-write the above inequality as $\langle f, (\lambda \cdot \Pi - \Pi A \Pi) f \rangle \geq 0$. However, note that the matrix $\lambda \cdot \Pi - \Pi A \Pi$ is actually a positive semidefinite matrix, which means that the expression $\langle f, (\lambda \cdot \Pi - \Pi A \Pi) f \rangle$ is a *sum-of-squares* of linear forms in the entries of f. The SoS hierarchy can be viewed as a proof system, where a solution to the level-$2t$ relaxation can be seen as satisfying all inequalities which can be derived using sum-of-squares of polynomials of degree at most t. We can show that this means that the solutions (after some modification) satisfy some codeword-like properties, using which it is possible to appeal to a *unique* decoding algorithm to recover one element from the list from one such "codeword-like" SoS solution. These codeword-like "distance certificates" for SoS solutions are developed in Section 5.

Broadly speaking, our techniques can be seen as part of the "Proofs to Algorithms" paradigm based on the Sum-of-Squares method [FKP19]. Our covering lemmas for SoS relaxations yield a generic framework for converting SoS proofs of distance for *any code*, to list decoding algorithms which work up to the Johnson bound. Examples for such results also include the earlier results by Richelson and Roy [RR22] for list decoding Ta-Shma's codes up to Johnson bound.

We also note that list-decoding algorithms often need to rely on algebraic structure, and are thus particularly well suited to work with large alphabets (fields). One then obtains algorithms for small-alphabet codes via techniques such as concatenation and list recovery. On the other hand, the techniques based on convex relaxations discussed above seem to work well directly over small alphabets.

1.4 Related work

In terms of techniques, the works most directly related to ours are those using similar SoS relaxations for list-decoding of Ta-Shma's codes [AJQ+20, JQST20, RR22]. In particular, the proofs of the covering lemmas follow the approach of Alev et al. [AJQ+20], and idea of viewing the proof of distance as implementable in the SoS hierarchy was also used by Richelson and Roy [RR22]. A precursor to much of this research on list-decoding, is the result of Dinur et al. [DHK+19], which suggested the approach of using semidefinite programming and expansion for list-decoding of codes obtained via an earlier distance amplification procedure of Alon et al. [ABN+92], which can be seen as a special case of the AEL distance amplification.

Another important work, related to the use of SoS hierarchy for decoding LDPC codes, is the *lower bound* of Ghazi and Lee [GL18] for using the SoS hierarchy to decode random LDPC codes. However, the lower bound shows that the relaxation for finding the optimal (closest) codeword may have value much better than the true optimal codeword, when the decoding radius is larger than that of LP decoding, thus showing that the SoS relaxations may not be integral. On the other hand, the relaxations we use do not optimize for the closest codeword, but rather go through covering lemmas. A recent work of Chen et al. [CCLO23] also shows significantly improved distance bounds, and improved unique-decoding bounds for the expander codes of Sipser and Spielman [SS96]. Our results do not apply for the codes considered in their work in a black box fashion, since the analysis is based on lossless *vertex expansion*, for which we do not always know of spectral certificates.

Our work can also be seen as obtaining "sparse" analogues of the results of Gopalan, Guruswami, and Raghavendra [GGR09] for the list-decoding of tensor and interleaved codes, which can be viewed as replacing the bipartite expanders in Zémor and AEL constructions respectively, by a complete bipartite graph.

2   Preliminaries and Notation

For a bipartite graph $G = (L, R, E)$ where L is the set of left vertices, and R is the set of right vertices, we index the left set by ℓ and the right set by r. For a vertex $\ell \in L$, we denote the set of edges incident to it by $N_L(\ell)$ (left neighborhood), and the set of edges incident to $r \in R$ is denoted by $N_R(r)$ (right neighborhood). We use $\ell \sim r$ to denote that the vertex $\ell \in L$ is adjacent to the vertex $r \in R$, that is, $(\ell, r) \in E$.

Fix an arbitrary ordering of the edges. Then there are bijections between the sets E, $L \times [d]$, and $R \times [d]$, given by taking (ℓ, i) to be the ith edge incident on ℓ, and similarly for $R \times [d]$. Henceforth, we will implicitly assume such an ordering of the edges is fixed, and use the resulting bijections.

Definition 2.1. *Let $[q]$ be a finite alphabet and let $f, g \in [q]^n$. Then the (fractional) distance between f, g is defined as*

$$\Delta(f, g) = \mathop{\mathbb{E}}_{i \in [n]} \left[\mathbf{1}\{f_i \neq g_i\} \right].$$

Definition 2.2 (Code, distance and rate). *A code $\mathcal{C}$ of block length n, distance δ and rate ρ over the alphabet size q is a set $\mathcal{C} \subseteq [q]^n$ with the following properties*

(i) $\rho = \frac{\log_q |\mathcal{C}|}{n}$

(ii) $\delta = \min_{h_1, h_2 \in \mathcal{C}} \Delta(h_1, h_2)$

Such codes are succinctly represented as $[n, \delta, \rho]_q$. We say $\mathcal{C}$ is a linear code if $[q]$ can be identified with a finite field and $\mathcal{C}$ is a linear subspace of $[q]^n$.

Definition 2.3 (List of codewords). *For any $g \in [q]^n$, the list of codewords in $\mathcal{C}$ that are at a distance less than δ from g is denoted by $\mathcal{L}(g, \delta)$. That is $\mathcal{L}(g, \delta) = \{h \in \mathcal{C} \ : \ \Delta(h, g) < \delta\}$.*

2.1   Expander graphs

Definition 2.4 ((n, d, λ)-expander). *A d-regular bipartite graph $G(L, R, E)$ with $|L| = |R| = n$ is said to be an (n, d, λ)-expander if*

$$\sigma_2(A_G) \leq \lambda \cdot d$$

where A_G is the $L \times R$ biadjacency matrix, and $\sigma_2(A_G)$ is its second largest singular value.

Infinite families of (n, d, λ)-expanders, with growing n as d and λ are constant, can be derived based on double covers of Ramanujan graphs of [LPS88] as long as $\lambda \geq \frac{2\sqrt{d-1}}{d}$.

Lemma 2.5 (Expander Mixing Lemma). *Given an (n, d, λ)-expander $G = (L, R, E)$ and functions $f : L \to \mathbb{R}$ and $g : R \to \mathbb{R}$, the following well known property is a simple consequence of definition of (n, d, λ)-expanders:*

$$\left| \mathop{\mathbb{E}}_{\ell \sim r} \left[f(\ell) \cdot g(r) \right] - \mathop{\mathbb{E}}_{\ell} \left[f(\ell) \right] \cdot \mathop{\mathbb{E}}_{r} \left[g(r) \right] \right| \leq \lambda \cdot \left(\mathop{\mathbb{E}}_{\ell} \left[f^2(\ell) \right] \right)^{1/2} \cdot \left(\mathop{\mathbb{E}}_{r} \left[g^2(r) \right] \right)^{1/2}.$$

2.2   Graph based codes

Expander codes.   We recap the construction from [Zém01], building on ideas from [Tan81, SS96], which constructs an infinite family of good codes starting from any good (inner) linear code over a small fixed block length of rate larger than $1/2$. The code $\mathcal{C}_0$ is also referred to as the base code for $\mathcal{C}^{Tan}$.

Definition 2.6. *Given an inner linear code $\mathcal{C}_0$ on alphabet $[q]$ and block length d, and a d-regular graph $G(V, E)$, we define the Tanner code $\mathcal{C}^{Tan}$ as*

$$\mathcal{C}^{Tan} = \{h : E \to [q] \ : \ h|_{N(v)} \in \mathcal{C}_0, \forall v \in V\}$$

where $h|_S$ for $S \subseteq E$ denotes the restriction of h to the set of coordinates indexed by S. In this work, we will only use Tanner codes defined on bipartite graphs.

By using an infinite family of graphs with constant degree d, we get an infinite family of codes based on an inner code of block length d.

Theorem 2.7 ([SS96], [Zém01]). *Let the distance and rate of the inner code $\mathcal{C}_0$ be δ_0 and ρ_0 respectively, and those of associated $\mathcal{C}^{Tan}$ be δ and ρ. If G is an (n, d, λ)-expander, then $\delta \geq \delta_0 \cdot (\delta_0 - \lambda)$ and $\rho \geq 2\rho_0 - 1$.*

The codes above were shown to be linear time decodable up to their unique decoding radius (for the designed distance) in [Zém01] and [SR03].

Alon-Edmonds-Luby distance amplification The following distance amplification scheme was introduced in [ABN+92], [AEL95] and used by [GI05] to design linear time unique decodable near-MDS codes.

Definition 2.8 (Concatenated codes). *Fix an (n, d, λ)-expander $G(L, R, E)$. Let $\mathcal{C}_1$ be an $[n, \delta_1, r_1]_{q_1}$ code and let $\mathcal{C}_0$ be a $[d, \delta_0, r_0]_{q_0}$ code with $q_1 = |\mathcal{C}_0|$.*

We define the concatenation of $f \in [q_1]^L$ with $\mathcal{C}_0 : [q_1] \to [q_0]^d$ as $f^* : E \to [q_0]$, defined as

$$f_{\mathcal{C}_0}^*(e) \ = \ \mathcal{C}_0(f(\ell))(j)$$

where ℓ is the left endpoint of edge e and e is the jth edge incident on ℓ. The concatenated code $\mathcal{C}_{\mathcal{C}_0}^*(\mathcal{C}_1)$ of block length $n \cdot d$ and alphabet $[q_0]$ is defined to be,

$$\mathcal{C}_{\mathcal{C}_0}^*(\mathcal{C}_1) \ = \ \{f_{\mathcal{C}_0}^* \ : \ f \in \mathcal{C}_1\}$$

When clear from context, we will omit $\mathcal{C}_0, \mathcal{C}_1$ in the above notation to call the concatenated code $\mathcal{C}^*$.

Claim 2.9. $\Delta(f_{\mathcal{C}_0}^*, g_{\mathcal{C}_0}^*) \ \geq \ \delta_0 \cdot \Delta(f, g)$, *which also implies* $\Delta(\mathcal{C}_{\mathcal{C}_0}^*(\mathcal{C}_1)) \geq \delta_0 \cdot \delta_1$.

Definition 2.10 (AEL Codes). *Fix an (n, d, λ)-expander $G(L, R, E)$. Let $\mathcal{C}_1$ be an $[n, \delta_1, r_1]_{q_1}$ code and let $\mathcal{C}_0$ be a $[d, \delta_0, r_0]_{q_0}$ code with $q_1 = |\mathcal{C}_0|$. We define the AEL-encoding $f_{\mathcal{C}_0}^{AEL} : R \to [q_0]^d$ as*

$$f_{\mathcal{C}_0}^{AEL}(r) \ = \ \left(f_{\mathcal{C}_0}^*(e_1), f_{\mathcal{C}_0}^*(e_2), \cdots, f_{\mathcal{C}_0}^*(e_d) \right)$$

where $e_1, e_2, \cdots, e_d$ are the d edges incident on r. The AEL code $\mathcal{C}_{\mathcal{C}_0}^{AEL}(\mathcal{C}_1) \subseteq [q_0^d]^n$ is defined as

$$\mathcal{C}_{\mathcal{C}_0}^{AEL}(\mathcal{C}_1) \ = \ \{f_{\mathcal{C}_0}^{AEL} \ : \ f \in \mathcal{C}_1\}$$

When clear from context, we will omit $\mathcal{C}_0, \mathcal{C}_1$ in the above notation to call the AEL code $\mathcal{C}^{AEL}$.

Alon, Edmonds and Luby, proved the following result, which shows that the construction can be used to amplify the distance to δ_0, by choosing λ sufficiently small.

Theorem 2.11 ([AEL95]). $\Delta(f_{\mathcal{C}_0}^{AEL}, g_{\mathcal{C}_0}^{AEL}) \ \geq \ \delta_0 - \frac{\lambda}{\Delta(f,g)}$, *which also implies* $\Delta(\mathcal{C}_{\mathcal{C}_0}^{AEL}(\mathcal{C}_1)) \ \geq \ \delta_0 - \frac{\lambda}{\delta_1}$.

A codeword f of $\mathcal{C}^{AEL}$ technically belongs to the space $[q_0^d]^R$. However, in this paper, we choose to identify codewords of $\mathcal{C}^{AEL}$ as belonging to $[q_0]^E$. It is clear that the two spaces are in bijection with each other and in fact, one can just *fold* or *unfold* the symbols to move from one space to the

other. Choosing f to be in $[q_0]^E$ allows us to talk about f viewed from left vertex set L or right vertex set R simply by changing the distance functions. Let $f, g \in \mathcal{C}^{AEL}$, then we define the following three distance functions:

$$\Delta^L(f,g) := \mathop{\mathbb{E}}_{\ell \in L} \left[\mathbf{1}\{f_{N_L(\ell)} \neq g_{N_L(\ell)}\} \right]$$

$$\Delta(f,g) := \mathop{\mathbb{E}}_{e \in E} \left[\mathbf{1}\{f_e \neq g_e\} \right] = \mathop{\mathbb{E}}_{\ell \in L} \left[\Delta(f_{N_L(\ell)}, g_{N_L(\ell)}) \right] = \mathop{\mathbb{E}}_{r \in R} \left[\Delta(f_{N_R(r)}, g_{N_R(r)}) \right]$$

$$\Delta^R(f,g) := \mathop{\mathbb{E}}_{r \in R} \left[\mathbf{1}\{f_{N_R(r)} \neq g_{N_R(r)}\} \right]$$

Theorem 2.11 can then be stated in a simpler form as

Theorem 2.12 (Restatement of Theorem 2.11)**.** $\Delta^R(f,g) \geq \delta_0 - \frac{\lambda}{\Delta^L(f,g)}$.

2.3 Sum-of-Squares hierarchy

The sum-of-squares hierarchy of semidefinite programs (SDPs) provides a family of increasingly powerful convex relaxations for several optimization problems. Each "level" t of the hierarchy is parametrized by a set of constraints corresponding to polynomials of degree at most t in the optimization variables. While the relaxations in the hierarchy can be viewed as semidefinite programs of size $n^{O(t)}$ [BS14, Lau09], it is often convenient to view the solution as a linear operator, called the "pseudoexpectation" operator.

Pseudoexpectations Let t be an positive even integer and fix an alphabet $[q]$. An SoS solution of degree t, or a pseudoexpectation of SoS-degree t, over the variables $\mathbf{Z} = \{Z_{i,j}\}_{i \in [m], j \in [q]}$ is represented by a linear operator $\widetilde{\mathbb{E}}[\cdot] : \mathbb{R}[\mathbf{Z}]^{\leq t} \to \mathbb{R}$ such that:

(i) $\widetilde{\mathbb{E}}[1] = 1$.

(ii) $\widetilde{\mathbb{E}}[p^2] \geq 0$ if p is a polynomial in $\mathbf{Z} = \{Z_{i,j}\}_{i \in [m], j \in [q]}$ of degree $\leq t/2$.

Note that linearity implies $\widetilde{\mathbb{E}}[p_1] + \widetilde{\mathbb{E}}[p_2] = \widetilde{\mathbb{E}}[p_1 + p_2]$ and $\widetilde{\mathbb{E}}[c \cdot p_1] = c \cdot \widetilde{\mathbb{E}}[p_1]$ for $c \in \mathbb{R}$, for $p_1, p_2 \in \mathbb{R}[\mathbf{Z}]^{\leq t}$. This also allows for a succinct representation of $\widetilde{\mathbb{E}}[\cdot]$ using any basis for $\mathbb{R}[\mathbf{Z}]^{\leq t}$.

The set of all pseudoexpectations should be seen as a relaxation for the set of all possible (distributions over) assignments to m variables in alphabet $[q]$. Indeed, any assignment $f : [m] \to [q]$, can be seen as a pseudoexpectation which assigns the value 1 to a monomial consistent with f and 0 otherwise. This can be extended via linearity to all polynomials, and then by convexity of the constraints to all distributions over assignments. However, the reverse is not true when $t < m$, and there can be degree-t pseudoexpectations which do not correspond to any genuine distribution.

It is possible to optimize over the set of degree-t pseudoexpectations in time $m^{O(t)}$ via SDPs (under certain conditions on the bit-complexity of solutions [O'D17, RW17]). We next define what it means for pseudoexpectations to satisfy some problem-specific constraints.

Definition 2.13 (Constrained Pseudoexpectations)**.** *Let $\mathcal{S} = \{f_1 = 0, \ldots, f_m = 0, g_1 \geq 0, \ldots, g_r \geq 0\}$ be a system of polynomial constraints, with each polynomial in $\mathcal{S}$ of degree at most t. We say $\widetilde{\mathbb{E}}[\cdot]$ is a pseudoexpectation operator respecting $\mathcal{S}$, if in addition to the above conditions, it also satisfies*

1. *$\widetilde{\mathbb{E}}[p \cdot f_i] = 0, \forall i \in [m]$ and $\forall p$ such that $\deg(p \cdot f_i) \leq t$.*

2. *$\widetilde{\mathbb{E}}[p^2 \cdot \prod_{i \in S} g_i] \geq 0, \forall S \subseteq [r]$ and $\forall p$ such that $\deg(p^2 \cdot \prod_{i \in S} g_i) \leq t$.*

Local constraints and local functions. Any constraint that involves at most k variables from $\mathbf{Z}$, with $k \le t$, can be written as a degree-k polynomial, and such constraints may be enforced into the SoS solution. In particular, we will always consider the following canonical constraints on the variables $\mathbf{Z}$.

$$Z_{i,j}^2 = Z_{i,j}, \ \forall i \in [m], j \in [q] \quad \text{and} \quad \sum_j Z_{i,j} = 1, \ \forall i \in [m].$$

We will also consider additional constraints and corresponding polynomials, defined by "local" functions. For any $f \in [q]^m$ and $M \subseteq [m]$, we use f_M to denote the restriction $f|_M$, and f_i to denote $f_{\{i\}}$ for convenience.

Definition 2.14 (*k-local function*). *A function $\mu : [q]^m \to \mathbb{R}$ is called k-local if there is a set $M \subseteq [m]$ of size k such that $\mu(f)$ only depends on $\{f(i)\}_{i \in M}$, or equivalently, $\mu(f)$ only depends on $f|_M$.*

If μ is k-local, we abuse notation to also use $\mu : [q]^M \to \mathbb{R}$ with $\mu(\alpha) = \mu(f)$ for any f such that $f|_M = \alpha$. It will be clear from the input to the function μ whether we are using μ as a function on $[q]^m$ or $[q]^M$.

Let $\mu : [q]^m \to \mathbb{R}$ be a k-local function that depends on coordinates $M \subseteq [m]$ with $|M| = k$. Then μ can be written as a degree-k polynomial p_μ in $\mathbf{Z}$:

$$p_\mu(\mathbf{Z}) = \sum_{\alpha \in [q]^M} \left(\mu(\alpha) \cdot \prod_{i \in M} Z_{i,\alpha_i} \right)$$

With some abuse of notation, we let $\mu(\mathbf{Z})$ denote $p_\mu(\mathbf{Z})$. We will use such k-local functions inside $\widetilde{\mathbb{E}}[\cdot]$ freely without worrying about their polynomial representation. For example, $\widetilde{\mathbb{E}}[\mathbf{1}\{\mathbf{Z}_i \ne j\}]$ denotes $\widetilde{\mathbb{E}}[1 - Z_{i,j}]$. The notion of k-local functions can also be extended from real valued functions to vector valued functions in a straightforward way.

Definition 2.15 (*vector-valued local functions*). *A function $\mu : [q]^m \to \mathbb{R}^N$ is k-local if the N real valued functions corresponding to the N coordinates are all k-local. Note that these different coordinate functions may depend on different sets of variables, as long as the number is at most k for each of the functions.*

Local distribution view of SoS It will be convenient to use a shorthand for the function $\mathbf{1}\{\mathbf{Z}_S = \alpha\}$, and we will use $\mathbf{Z}_{S,\alpha}$. Likewise, we use $\mathbf{Z}_{i,j}$ as a shorthand for the function $\mathbf{1}\{\mathbf{Z}_i = j\}$. That is, henceforth,

$$\widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}\right] = \widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_S = \alpha\}\right] = \widetilde{\mathbb{E}}\left[\prod_{s \in S} Z_{s,\alpha_s}\right] \quad \text{and} \quad \widetilde{\mathbb{E}}\left[\mathbf{Z}_{i,j}\right] = \widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_i = j\}\right] = \widetilde{\mathbb{E}}\left[Z_{i,j}\right]$$

Note that for any $S \subseteq [m]$ with $|S| = k \le t/2$,

$$\sum_{\alpha \in [q]^k} \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}\right] = \widetilde{\mathbb{E}}\left[\prod_{s \in S} \left(\sum_{j \in [q]} Z_{s,j}\right)\right] = 1 \quad \text{and} \quad \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}\right] = \widetilde{\mathbb{E}}\left[\prod_{s \in S} Z_{s,\alpha_s}\right] = \widetilde{\mathbb{E}}\left[\prod_{s \in S} Z_{s,\alpha_s}^2\right] \ge 0.$$

Thus, the values $\left\{\widetilde{\mathbb{E}}[\mathbf{Z}_{S,\alpha}]\right\}_{\alpha \in [q]^S}$ define a distribution over $[q]^k$, referred to as the local distribution for $\mathbf{Z}_S$.

Let $\mu : [q]^m \to \mathbb{R}$ be a k-local function for $k \le t/2$, depending on $M \subseteq [m]$. Then, $\widetilde{\mathbb{E}}[\mu(\mathbf{Z})]$ can be seen as the expected value of the function μ under the local distribution for M, since

$$\widetilde{\mathbb{E}}\left[\mu(\mathbf{Z})\right] = \widetilde{\mathbb{E}}\left[\sum_{\alpha \in [q]^M} \left(\mu(\alpha) \cdot \prod_{i \in M} Z_{i,\alpha_i}\right)\right] = \sum_{\alpha \in [q]^M} \mu(\alpha) \cdot \widetilde{\mathbb{E}}\left[\prod_{i \in M} Z_{i,\alpha_i}\right] = \sum_{\alpha \in [q]^M} \mu(\alpha) \cdot \widetilde{\mathbb{E}}\left[\mathbf{Z}_{M,\alpha}\right].$$

8

Claim 2.16. *Let $\widetilde{\mathbb{E}}[\cdot]$ be a degree-t pseudoexpectation. For $k \leq t/2$, let μ_1, μ_2 be two k-local functions on $[q]^m$, depending on the same set of coordinates M, and $\mu_1(\alpha) \leq \mu_2(\alpha) \ \forall \alpha \in [q]^M$. Then $\widetilde{\mathbb{E}}[\mu_1(\mathbf{Z})] \leq \widetilde{\mathbb{E}}[\mu_2(\mathbf{Z})]$.*

Proof. Let $\mathcal{D}_M$ be the local distribution induced by $\widetilde{\mathbb{E}}[\cdot]$ for $\mathbf{Z}_M$. Then $\widetilde{\mathbb{E}}[\mu_1(\mathbf{Z})] = \mathbb{E}_{\alpha \sim \mathcal{D}_M}[\mu_1(\alpha)]$, and $\widetilde{\mathbb{E}}[\mu_2(\mathbf{Z})] = \mathbb{E}_{\alpha \sim \mathcal{D}_M}[\mu_2(\alpha)]$, which implies $\widetilde{\mathbb{E}}[\mu_1(\mathbf{Z})] \leq \widetilde{\mathbb{E}}[\mu_2(\mathbf{Z})]$. $\blacksquare$

The previous claim allows us to replace any local function inside $\widetilde{\mathbb{E}}[\cdot]$ by another local function that dominates it. We will make extensive use of this fact.

Covariance for SoS solutions Given two sets $S, T \subseteq [m]$ with $|S|, |T| \leq k/4$, we can define the covariance between indicator random variables of $\mathbf{Z}_S$ and $\mathbf{Z}_T$ taking values α and β respectively, according to the local distribution over $S \cup T$. This is formalized in the next definition.

Definition 2.17. *Let $\widetilde{\mathbb{E}}[\cdot]$ be a pseudodistribution operator of SoS-degree-t, and S, T are two sets of size at most $t/4$, and $\alpha \in [q]^S$, $\beta \in [q]^T$, we define the pseudo-covariance and pseudo-variance,*

$$\widetilde{\mathrm{Cov}}(\mathbf{Z}_{S,\alpha}, \mathbf{Z}_{T,\beta}) \ = \ \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha} \cdot \mathbf{Z}_{T,\beta}\right] - \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}\right]\widetilde{\mathbb{E}}\left[\mathbf{Z}_{T,\beta}\right] \qquad and \qquad \widetilde{\mathrm{Var}}\left[\mathbf{Z}_{S,\alpha}\right] \ = \ \widetilde{\mathrm{Cov}}(\mathbf{Z}_{S,\alpha}, \mathbf{Z}_{S,\alpha})$$

The above definition is extended to pseudo-covariance and pseudo-variance for pairs of sets S, T, as the sum of absolute value of pseudo-covariance for all pairs α, β :

$$\widetilde{\mathrm{Cov}}(\mathbf{Z}_S, \mathbf{Z}_T) \ = \ \sum_{\substack{\alpha \in [q]^S \\ \beta \in [q]^T}} \left|\widetilde{\mathrm{Cov}}(\mathbf{Z}_{S,\alpha}, \mathbf{Z}_{T,\beta})\right| \qquad and \qquad \widetilde{\mathrm{Var}}\left[\mathbf{Z}_S\right] \ = \ \sum_{\alpha \in [q]^S} \left|\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{S,\alpha}\right]\right|$$

We will need the fact that $\widetilde{\mathrm{Var}}[\mathbf{Z}_S]$ is bounded above by 1, since,

$$\widetilde{\mathrm{Var}}\left[\mathbf{Z}_S\right] \ = \ \sum_\alpha \left|\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{S,\alpha}\right]\right| \ = \ \sum_\alpha \left(\widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}^2\right] - \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}\right]^2\right) \ \leq \ \sum_\alpha \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}^2\right] \ = \ \sum_\alpha \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}\right] \ = \ 1$$

Conditioning SoS solutions. We will also make use of conditioned pseudoexpectation operators, which may be defined in a way similar to usual conditioning for true expectation operators, as long as the event we condition on is local. The conditioned SoS solution is of a smaller degree, but continues to respect the constraints that original solution respects.

Definition 2.18. *Let $F \subseteq [q]^m$ be subset (to be thought of as an event) such that $\mathbf{1}_F : [q]^m \to \{0, 1\}$ is a k-local function. Then for every $t > 2k$, we can condition a pseudoexpectation operator of SoS-degree t on F to obtain a new conditioned pseudoexpectation operator $\widetilde{\mathbb{E}}[\cdot \mid E]$ of SoS-degree $t - 2k$, as long as $\widetilde{\mathbb{E}}[\mathbf{1}_F^2(\mathbf{Z})] > 0$. The conditioned SoS solution is given by*

$$\widetilde{\mathbb{E}}\left[p(\mathbf{Z}) \mid F(\mathbf{Z})\right] := \frac{\widetilde{\mathbb{E}}[p(\mathbf{Z}) \cdot \mathbf{1}_F^2(\mathbf{Z})]}{\widetilde{\mathbb{E}}[\mathbf{1}_F^2(\mathbf{Z})]}$$

where p is any polynomial of degree at most $t - 2k$.

We can also define pseudocovariances and pseudo-variances for the conditioned SoS solutions.

Definition 2.19. *Let $F \subseteq [q]^m$ be an event such that $\mathbf{1}_F$ is k-local, and let $\widetilde{\mathbb{E}}[\cdot]$ be a pseudoexpectation operator of degree t, with $t > 2k$. Let S, T be two sets of size at most $\frac{t-2k}{2}$ each. Then the pseudocovariance between $\mathbf{Z}_{S,\alpha}$ and $\mathbf{Z}_{T,\beta}$ for the solution conditioned on event F is defined as,*

$$\widetilde{\mathrm{Cov}}(\mathbf{Z}_{S,\alpha}, \mathbf{Z}_{T,\beta}|F) \ = \ \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}\mathbf{Z}_{T,\beta}|F\right] - \widetilde{\mathbb{E}}\left[\mathbf{Z}_{S,\alpha}|F\right]\widetilde{\mathbb{E}}\left[\mathbf{Z}_{T,\beta}|F\right]$$

We also define the pseudocovariance between $\mathbf{Z}_{S,\alpha}$ and $\mathbf{Z}_{T,\beta}$ after conditioning on a random assignment for some $\mathbf{Z}_V$ with $V \subseteq [m]$. Note that here the random assignment for $\mathbf{Z}_V$ is chosen according to the local distribution for the set V.

Definition 2.20 (Pseudocovariance for conditioned pseudoexpectation operators).

$$\widetilde{\mathrm{Cov}}(\mathbf{Z}_{S,\alpha}, \mathbf{Z}_{T,\beta} | \mathbf{Z}_V) = \sum_{\gamma \in [q]^V} \widetilde{\mathrm{Cov}}(\mathbf{Z}_{S,\alpha}, \mathbf{Z}_{T,\beta} | \mathbf{Z}_V = \gamma) \cdot \widetilde{\mathbb{E}}\left[\mathbf{Z}_{V,\gamma}\right]$$

And we likewise define $\widetilde{\mathrm{Var}}[\mathbf{Z}_{S,\alpha} | \mathbf{Z}_V]$, $\widetilde{\mathrm{Cov}}(\mathbf{Z}_S, \mathbf{Z}_T | \mathbf{Z}_V)$ and $\widetilde{\mathrm{Var}}[\mathbf{Z}_S | \mathbf{Z}_V]$.

2.4 SoS relaxations for codes

For both Tanner codes and AEL codes, we will identify $[m]$ with E so that the SoS solutions will be relaxations to the assignments to edges of a bipartite (n, d, λ)-expander.

Pseudocodewords for Tanner Codes   Let $G(L, R, E)$ be the bipartite (n, d, λ)-expander on which the Tanner code is defined, and let $\mathcal{C}_0 \subseteq [q]^d$ be the inner code. The SoS variables will be $\mathbf{Z} = \{Z_{e,j}\}_{e \in E, j \in [q]}$.

Definition 2.21 (Tanner Pseudocodewords). *For $t \geq 2d$, we define a degree-t Tanner pseuocodeword to be a degree-t pseudoexpectation operator $\widetilde{\mathbb{E}}[\cdot]$ on $\mathbf{Z}$ respecting the following constraints:*

$$\forall \ell \in L, \quad \mathbf{Z}_{N_L(\ell)} \in \mathcal{C}_0 \qquad and \qquad \forall r \in R, \quad \mathbf{Z}_{N_R(r)} \in \mathcal{C}_0$$

Again, since these constraints are d-local, it is sufficient to enforce that certain degree-d polynomials are zero (respected by pseudoexpectation) to enforce these constraints. In particular, each parity check in the parity check matrix of $\mathcal{C}_0$ will correspond to a monomial of size at most d that we can enforce to be equal to 1.

We can also define a generalization of distance between two codewords to include pseudocodewords.

Definition 2.22 (Distance from a pseudocodeword). *The distance between a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ of SoS-degree $t \geq 2d$ and a codeword h of $\mathcal{C}^{Tan}$ is defined as*

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) := \mathop{\mathbb{E}}_{e \in E}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_e \neq h_e\}\right]\right] = \mathop{\mathbb{E}}_{\ell \in L}\left[\widetilde{\mathbb{E}}\left[\Delta(\mathbf{Z}_{N_L(\ell)}, h_{N_L(\ell)})\right]\right]$$

Pseudocodewords for AEL   Let $G(L, R, E)$ be the bipartite (n, d, λ)-expander on which the AEL code is defined, and let $\mathcal{C}_0 \subseteq [q_0]^d$ be the inner code. The SoS variables will be $\mathbf{Z} = \{Z_{e,j}\}_{e \in E, j \in [q_0]}$.

Definition 2.23 (AEL Pseudocodewords). *For $t \geq 2d$, we define a degree-t AEL pseuocodeword to be a degree-t pseudoexpectation operator $\widetilde{\mathbb{E}}[\cdot]$ on $\mathbf{Z}$ respecting the following constraints:*

$$\forall \ell \in L, \quad \mathbf{Z}_{N_L(\ell)} \in \mathcal{C}_0$$

Next we define the distances between a pseudocodeword and a codeword of $\mathcal{C}^{AEL}$.

Definition 2.24 (Distance from a pseudocodeword). *The left, middle and right distances between a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ of SoS-degree $t \geq 2d$ and a codeword $h \in \mathcal{C}^{AEL}$ are defined as*

$$\Delta^L(\widetilde{\mathbb{E}}[\cdot], h) := \mathop{\mathbb{E}}_{\ell}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\right]\right]$$

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) := \mathop{\mathbb{E}}_{e}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_e \neq h_e\}\right]\right]$$

$$\Delta^R(\widetilde{\mathbb{E}}[\cdot], h) := \mathop{\mathbb{E}}_{r \in R}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}\right]\right]$$

3 Proof Overview

Our proof can be viewed as an algorithmic implementation of the proof of Johnson bound, and the proofs of distance, for the relevant codes. We start with an overview of the proof for Tanner codes. The argument is very similar for all codes considered here, substituting an appropriate proof of distance in each case.

Johnson bound via covering lemmas. We first prove the Johnson bound via a statement we will call a "covering lemma", which can then be generalized to work with convex relaxations. Given an $[m, \delta, \rho]_q$ code $\mathcal{C}$ and a received word $g \in [q]^m$, our goal is to bound the list of radius $\mathcal{L}(g, \mathcal{J}_q(\delta)) = \{h \in \mathcal{C} : \Delta(h, g) < \mathcal{J}_q(\delta)\}$, where $\mathcal{J}_q(\delta)$ denotes the Johnson bound. It will be convenient to work with $\beta = 1 - \frac{q \cdot \delta}{q-1}$, and the Johnson bound can be defined using the equation

$$\left(1 - \frac{q \cdot \mathcal{J}_q(\delta)}{q-1}\right)^2 = \left(1 - \frac{q \cdot \delta}{q-1}\right) = \beta.$$

Also, we can map elements of $[q]$ to corners $u_1, \ldots, u_q$ of the simplex in $\mathbb{R}^{q-1}$, which are unit vectors satisfying $\langle u_i, u_j \rangle = -1/(q-1)$ for $i \neq j$. Applying this map, say χ, pointwise to $g, h \in [q]^m$, gives

$$\langle \chi(g), \chi(h) \rangle = \mathbb{E}_{i \in [m]} [\langle \chi(g(i)), \chi(h(i)) \rangle] = 1 - \Delta(g, h) - \frac{\Delta(g, h)}{q-1} = 1 - \frac{q \cdot \Delta(g, h)}{q-1}.$$

Thus, given g and β as above, we can write the list as $\mathcal{L} = \{h \in \mathcal{C} : \langle \chi(g), \chi(h) \rangle > \sqrt{\beta}\}$. The covering lemma in Section 4 shows that given a set $\mathcal{F}$ of unit vectors in an inner-product space, and a unit vector $\tilde{g}$ satisfying $\langle \tilde{g}, f \rangle > \sqrt{\beta}$ for all $f \in \mathcal{F}$, there exists g_0 in the *convex hull* $\operatorname{conv}(\mathcal{F})$ satisfying $\langle g_0, f \rangle > \beta$ for all $f \in \mathcal{F}$.

Instantiating this with $\tilde{g} = \chi(g)$ and $\mathcal{F} = \{\chi(h) : h \in \mathcal{L}\}$ gives $g_0 \in \operatorname{conv}(\mathcal{F})$ satisfying $\langle g_0, \chi(h) \rangle > \beta$ for all $h \in \mathcal{L}$. Additionally, g_0 can be chosen to be supported on at most $m \cdot (q-1) + 1$ elements of $\mathcal{F}$ via Carathéodory theorem. Since $\langle \chi(h_1), \chi(h_2) \rangle \leq \beta$ for any $h_1 \neq h_2$ in $\mathcal{C}$, any $h \in \mathcal{F} \setminus \operatorname{Supp}(g_0)$ will satisfy $\langle g_0, \chi(h) \rangle \leq \beta$, which is a contradiction. Thus, we must have $\mathcal{L} \subseteq \operatorname{Supp}(g_0)$ implying $|\mathcal{L}| \leq (q-1) \cdot m + 1$. While this is a weaker bound on the list size (but can be independent of m via approximate Carathéodory theorems), each step of the above proof can be extended to work well with convex relaxations.

Algorithmic covering lemmas via SoS. Note that in the above proof, it suffices to have g_0 in the convex hull of *all* codewords i.e., $g_0 \in \operatorname{conv}(\chi(h) : h \in \mathcal{C})$, instead of only the codewords in $\mathcal{L}$. Since the convex hull of all the codewords is still a difficult set to optimize over, we instead work with degree-t *pseudocodewords* defined as solutions to an SoS relaxation of degree-t, respecting certain local constraints corresponding to the code (see Section 2).

Moreover, the covering lemma used above can be proved by finding the $g_0 \in \operatorname{conv}(\mathcal{F})$, which minimizes $\|g_0\|$ while satisfying $\langle g_0, g \rangle > \sqrt{\beta}$. Analogously, we consider solutions to the SoS relaxation, given as pseudocodewords $\widetilde{\mathbb{E}}[\cdot]$ satisfying $\left\langle \widetilde{\mathbb{E}}[\chi(\mathbf{Z})], \chi(g) \right\rangle > \sqrt{\beta}$ and minimizing $\left\|\widetilde{\mathbb{E}}[\chi(\mathbf{Z})]\right\|$. Note that here $\chi(\mathbf{Z})$ is a (vector-valued) 1-local function, and thus the vector $\widetilde{\mathbb{E}}[\chi(\mathbf{Z})]$ is well defined. A similar relaxation was also used for list-decoding of direct-sum codes in [AJQ+20].

Choosing from the "support" via conditioning. The next step of the proof of Johnson bound can be seen as saying that for all h_1 such that $\chi(h_1) \in \operatorname{Supp}(g_0)$ and $h_2 \in \mathcal{L}$, we must have $\langle \chi(h_1), \chi(h_2) \rangle = 1$ (when they are equal) or $\langle \chi(h_1), \chi(h_2) \rangle \leq \beta$ (when they are distinct codewords).

We will do this in two steps. The first is to develop a good proxy for "$\chi(h_1) \in \operatorname{Supp}(g_0)$" since we are working with the pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$, which is not a convex combination of codewords.

11

Instead, it suffices to look at pseudocodewords which have small pseudo-covariance across a typical pair of left-right vertices in the bipartite graph defining the Tanner code i.e.,

$$\mathop{\mathbb{E}}_{\substack{\ell \in L \\ r \in R}} \left[\widetilde{\mathrm{Cov}}(\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)}) \right] \leq \eta.$$

Note that when $\widetilde{\mathbb{E}}[\cdot]$ corresponds to an actual codeword (or any integral solution) the covariance will be 0. Thus, the above is a weakening of the notion of "vertex of the convex hull". We call such a solution, an η-*good* pseudocodeword. A similar definition was also used by Richelson and Roy [RR22] in their list-decoding algorithm for Ta-Shma's codes [TS17], and also in earlier works [AJQ+20, JQST20].

An argument of Barak, Raghavendra, and Steurer [BRS11] shows that conditioning the starting SoS solution on the values few randomly chosen variables, leads to an η-good solution. In fact, one either has small covariance in the sense above, or conditioning on the variables in $N_R(r)$ for a random $r \in R$ reduces the average variance (seen from the left) by $\Omega_{q,d}(\eta^2)$. Since the (pseudo-)variance is non-negative, this process terminates, yielding an η-good solution. The argument can also be made deterministic by enumerating over all (constant-sized) subsets to condition on.

Proof of distance for good pseudocodewords. Given an η-good pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ as defined above, we can now prove that for any $h \in \mathcal{L}$, we must have

$$\left\langle \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z}) \right], \chi(h) \right\rangle \geq 1 - O(\eta) \qquad \text{or} \qquad \left\langle \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z}) \right], \chi(h) \right\rangle \leq \beta + O(\eta).$$

To argue the above dichotomy, we will switch to the distances instead of inner products, as the proof closely follows the distance proof of Tanner codes. Let $\{X_\ell\}_{\ell \in L}$ and $\{Y_r\}_{r \in R}$ be ensembles of d-local functions defined as $X_\ell(\mathbf{Z}) = \mathbf{1}\left\{ \mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)} \right\}$ and $Y_r(\mathbf{Z}) = \mathbf{1}\left\{ \mathbf{Z}_{N_R(r)} \neq h_{N_R(r)} \right\}$. Using the fact that the distance of the base code $\mathcal{C}_0$ is at least δ_0, and the fact that the local constraints for $\mathcal{C}_0$ are respected by the pseudocodewords, one can show that

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) := \mathop{\mathbb{E}}_{e \in E}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\left\{ \mathbf{Z}_e \neq h_e \right\} \right] \right] \geq \delta_0 \cdot \mathop{\mathbb{E}}_{\ell \in L}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z}) \right] \right].$$

Taking the geometric mean with a similar inequality for $\{Y_r\}_{r \in R}$ gives

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) \geq \delta_0 \cdot \left(\mathop{\mathbb{E}}_{\substack{\ell \in L \\ r \in R}}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z}) \right] \cdot \widetilde{\mathbb{E}}\left[Y_r(\mathbf{Z}) \right] \right] \right)^{1/2}.$$

On the other hand, a variant of the expander mixing lemma for pseudoexpectations (see Section 5), together with the η-good property and the simple observation that

$$\mathbf{1}\{\mathbf{Z}_e \neq h_e\} \leq X_\ell(\mathbf{Z}) \cdot Y_r(\mathbf{Z}),$$

gives

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) \leq \mathop{\mathbb{E}}_{\ell \sim r}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z}) \cdot Y_r(\mathbf{Z}) \right] \right]$$

$$\leq \mathop{\mathbb{E}}_{\substack{\ell \in L \\ r \in R}}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z}) \cdot Y_r(\mathbf{Z}) \right] \right] + \lambda \cdot \left(\mathop{\mathbb{E}}_{\substack{\ell \in L \\ r \in R}}\left[\widetilde{\mathbb{E}}\left[X_\ell^2(\mathbf{Z}) \right] \cdot \widetilde{\mathbb{E}}\left[Y_r^2(\mathbf{Z}) \right] \right] \right)^{1/2}$$

$$\leq \mathop{\mathbb{E}}_{\substack{\ell \in L \\ r \in R}}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z}) \right] \cdot \widetilde{\mathbb{E}}\left[Y_r(\mathbf{Z}) \right] \right] + \eta + \lambda \cdot \left(\mathop{\mathbb{E}}_{\substack{\ell \in L \\ r \in R}}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z}) \right] \cdot \widetilde{\mathbb{E}}\left[Y_r(\mathbf{Z}) \right] \right] \right)^{1/2}.$$

Note that the last line also used $\widetilde{\mathbb{E}}[X_\ell^2(\mathbf{Z})] = \widetilde{\mathbb{E}}[X_\ell(\mathbf{Z})]$ (and similarly for Y_r). Comparing the two bounds and solving the resulting quadratic inequality in $\tau = (\mathbb{E}_{\ell,r}[\widetilde{\mathbb{E}}[X_\ell(\mathbf{Z})] \cdot \widetilde{\mathbb{E}}[Y_r(\mathbf{Z})]])^{1/2}$ yields the dichotomy.

12

Completing the argument. Starting the algorithmic covering lemma with parameter $\beta + \varepsilon$ ensures that with positive (constant) probability, for any fixed $h \in \mathcal{L}$, the conditioning yields an η-good pseudocodeword satisfying $\left\langle \widetilde{\mathbb{E}}[\chi(\mathbf{Z})], \chi(h) \right\rangle \geq \beta + \varepsilon/2$. By the above dichotomy, we must be in the first case (for sufficiently small η). A simple averaging argument then shows that considering $h' \in [q]^m$ with $h'(e) = \text{argmax}_{j \in [q]} \left\{ \widetilde{\mathbb{E}}[\mathbf{1}\{\mathbf{Z}_e = j\}] \right\}$ gives $\Delta(h', h) = O(\eta)$. Given such an h', the codeword h can be recovered by unique decoding.

The above argument can also be made to work for other codes, where the distance proof is a spectral argument based on expanders, such as the codes obtained via the distance amplification of Alon, Edmonds and Luby [AEL95]. One can simply use pseudocodewords for the appropriate code, and substitute the corresponding distance proof in the argument above.

4   Covering Lemma and Johnson Bounds

In this section, we will introduce our abstract covering lemma, and then use it to give algorithm-friendly proofs of (known) Johnson bounds [Gur01] by showing that there is a distribution over codewords that covers the list. Next, we will show that if we are willing to work with distributions over degree-t pseudocodewords (or just pseudocodewords, due to convexity), then such a pseudocodeword may be found in time $n^{\mathcal{O}(t)}$.

4.1   Covering Lemma

Lemma 4.1 (Covering Lemma). *Let $\mathcal{H}$ be an inner product space, and let $\mathcal{F}$ be a family of unit vectors in $\mathcal{H}$. Suppose there exists a $g \in \mathcal{H}$ of unit norm such that for any $f \in \mathcal{F}$, $\langle g, f \rangle > \varepsilon$. Then, there exists a $g_0 \in \text{conv}(\mathcal{F})$ such that for any $f \in \mathcal{F}$, $\langle g_0, f \rangle > \varepsilon^2$.*

Proof. Consider the set $T = \{v \in \text{conv}(\mathcal{F}) : \langle v, g \rangle > \varepsilon\}$, which is non-empty because any $f \in \mathcal{F}$ also belongs to T. Let $g_0 = \text{argmin}_{v \in T} \|v\|^2$. We will show that g_0 must have the property that $\langle g_0, f \rangle > \varepsilon^2$ for any $f \in \mathcal{F}$. Note that $\|g_0\| \cdot \|g\| \geq \langle g_0, g \rangle > \varepsilon$ means $\|g_0\| > \varepsilon$.

Suppose not, then there exists an $f \in \mathcal{F}$ such that $\langle g_0, f \rangle \leq \varepsilon^2$. For an $\theta \in [0, 1]$ to be chosen later, consider $g_1 = \theta g_0 + (1 - \theta) f$, which is also in $\text{conv}(\mathcal{F})$, and $\langle g_1, g \rangle = \theta \cdot \langle g_0, g \rangle + (1 - \theta) \cdot \langle f, g \rangle > \varepsilon$.

$$
\begin{aligned}
\|g_1\|^2 &= \langle \theta g_0 + (1 - \theta)f, \theta g_0 + (1 - \theta)f \rangle \\
&= \theta^2 \cdot \|g_0\|^2 + 2\theta(1 - \theta) \cdot \langle g_0, f \rangle + (1 - \theta)^2 \cdot \|f\|^2 \\
&\leq \theta^2 \cdot \|g_0\|^2 + 2\theta(1 - \theta)\varepsilon^2 + (1 - \theta)^2 \\
&= \theta^2 \cdot (\|g_0\|^2 - \varepsilon^2) + \varepsilon^2 \cdot (\theta + (1 - \theta))^2 + (1 - \theta)^2(1 - \varepsilon^2) \\
&= \theta^2 \cdot (\|g_0\|^2 - \varepsilon^2) + \varepsilon^2 + (1 - \theta)^2 \cdot (1 - \varepsilon^2) \\
&= \theta^2 \cdot (1 - \varepsilon^2 + \|g_0\|^2 - \varepsilon^2) - 2\theta \cdot (1 - \varepsilon^2) + 1
\end{aligned}
$$

At $\theta = 1$, this expression equals $\|g_0\|^2$. The minimum of this quadratic function of θ is achieved at $\theta = \frac{2(1 - \varepsilon^2)}{2(1 - \varepsilon^2 + \|g_0\|^2 - \varepsilon^2)}$, which is < 1 as $\|g_0\| > \varepsilon$. Thus, we can reduce $\|g_1\|^2$ further to strictly less than $\|g_0\|^2$ by choosing θ to the above value, which contradicts the optimality of g_0. ∎

4.2   Johnson bounds

We will prove the standard q-ary Johnson bound and the version from [GS01] which receives weight $W_{i,j}$ for each $j \in [q]$ for every coordinate $i \in [n]$. First we define some functions to embed the

13

received word g (or received weights) in $\mathbb{R}^{(q-1)n}$, which will be the inner product space where we apply the covering lemma.

Definition 4.2. *Fix $q \in \mathbb{N}$. We denote by $\chi_q : [q] \to \mathbb{R}^{q-1}$ any function that satisfies*

$$\langle \chi_q(j_1), \chi_q(j_2) \rangle = \begin{cases} 1 & j_1 = j_2 \\ \frac{-1}{q-1} & j_1 \neq j_2 \end{cases}$$

For $f \in [q]^n$, we use $\chi_q(f)$ to denote the vector in $\mathbb{R}^{(q-1)n}$ obtained by applying χ_q on f coordinate-wise. Note that

$$\|\chi_q(f)\|^2 = \mathop{\mathbb{E}}_{i \in [n]} \left[\|\chi_q(f_i)\|^2 \right] = 1$$

When clear from context, we may omit q to write $\chi(f)$.

Such a χ_q exists because the corresponding $q \times q$ Gram matrix is positive semidefinite and of rank $q - 1$.

Observation 4.3. *For $f_1, f_2 \in [q]^n$, if $\Delta(f_1, f_2) = (1 - \frac{1}{q})(1 - \beta)$, then*

$$\langle \chi(f), \chi(g) \rangle = \beta$$

Theorem 4.4 (q-ary Johnson bound). *Let $\mathcal{C} \subseteq [q]^n$ be a code of distance at least $\Delta(\mathcal{C}) = (1 - \frac{1}{q})(1 - \beta)$. For any $g \in [q]^n$,*

$$\left| \mathcal{L}\left(g, \left(1 - \frac{1}{q}\right) \cdot (1 - \sqrt{\beta}) \right) \right| \leq (q - 1) \cdot n + 1$$

Proof. Let $\mathcal{L} = \mathcal{L}\left(g, \left(1 - \frac{1}{q}\right) \cdot (1 - \sqrt{\beta}) \right) = \{h_1, h_2, \cdots, h_m\}$. Then for any $h_i \in \mathcal{L}$,

$$\langle \chi(g), \chi(h_i) \rangle > \sqrt{\beta}$$

By Lemma 4.1, there exists a $g_0 \in \mathrm{conv}\left(\{\chi(h_1), \chi(h_2), \cdots, \chi(h_m)\} \right)$ such that $\langle g_0, \chi(h) \rangle > \beta$ for every $h \in \mathcal{L}$. Since $g_0 \in \mathbb{R}^{(q-1)n}$, by the Carathéodory theorem, we may assume that g_0 can be written as a convex combination of at most $(q - 1)n + 1$ elements of $\{\chi(h_1), \chi(h_2), \cdots, \chi(h_m)\}$. Let this subset of $\{\chi(h_1), \chi(h_2), \cdots, \chi(h_m)\}$ be $\mathrm{Supp}(g_0)$.

For any $h \in \mathcal{L}$,

$$\begin{aligned} \langle g_0, \chi(h) \rangle > \beta &\Rightarrow \exists \chi(h_0) \in \mathrm{Supp}(g_0) \text{ such that } \langle \chi(h_0), \chi(h) \rangle > \beta \\ &\Rightarrow \exists \chi(h_0) \in \mathrm{Supp}(g_0) \text{ such that } \langle \chi(h_0), \chi(h) \rangle = 1 \\ &\Rightarrow \exists \chi(h_0) \in \mathrm{Supp}(g_0) \text{ such that } h_0 = h \\ &\Rightarrow \chi(h) \in \mathrm{Supp}(g_0) \end{aligned}$$

Here we used the fact that for any $h_i, h_j \in \mathcal{L}$, due to distance of the code,

$$\langle \chi(h_i), \chi(h_j) \rangle \leq \beta \quad \text{or} \quad \langle \chi(h_i), \chi(h_j) \rangle = 1$$

Finally, we have concluded that $\{\chi(h_1), \chi(h_2), \cdots, \chi(h_m)\} \subseteq \mathrm{Supp}(g_0)$ and using $|\mathrm{Supp}(g_0)| \leq (q - 1) \cdot n + 1$ gives $m = |\mathcal{L}| \leq (q - 1) \cdot n + 1$. ∎

Note that in the above proof, $g_0 \in \mathrm{conv}\left(\{\chi(h_1), \chi(h_2), \cdots, \chi(h_m)\} \right) \subseteq \mathrm{conv}\left(\chi(\mathcal{C}) \right)$. Indeed, even if we minimize the norm as in the proof of Lemma 4.1 over $\mathrm{conv}\left(\chi(\mathcal{C}) \right)$, we will still get the covering property. This will be useful when trying to find the cover via an efficient algorithm, where we do not know the list apriori.

We now prove the more general weighted version of Johnson bound, which captures list recovery as a special case.

Theorem 4.5 (Weighted Johnson bound [Gur01]). *Let distance of code be at least* $(1 - \frac{1}{q})(1 - \beta)$. *Given weights* $\{w_{i,j}\}_{i \in [n], j \in [q]}$, *let* $W_i = \sum_j w_{i,j}$ *and* $W_i^{(2)} = \sum_j w_{i,j}^2$. *The number of codewords h that satisfy*

$$\mathbb{E}_i \left[\frac{w_{i,h_i}}{W_i} \right] > \frac{1}{q} + \sqrt{\left(1 - \frac{1}{q}\right) \cdot \mathbb{E}_i \left[\frac{W_i^{(2)}}{W_i^2} - \frac{1}{q} \right] \cdot \beta}$$

is at most $n(q-1) + 1$.

Proof. Embed the received weights in i^{th} position as $\chi(\{w_{i,j}\}_{j \in [q]}) = \sum_j \frac{w_{i,j} \chi_q(j)}{W_i}$, and append all these n vectors, each of dimension $q - 1$, and normalize to a unit vector to form the final embedded vector u. This normalizing factor is $\left(\mathbb{E}_i \left\| \frac{\sum_j w_{ij} \chi_q(j)}{W_i} \right\|^2 \right)^{1/2}$, which we can simplify as

$$
\begin{aligned}
\mathbb{E}_i \left[\left\| \frac{\sum_j w_{ij} \chi_q(j)}{W_i} \right\|^2 \right] &= \mathbb{E}_i \left[\frac{1}{W_i^2} \left(\sum_j w_{ij}^2 - \frac{1}{q-1} \sum_{j_1 \neq j_2} w_{ij_1} w_{ij_2} \right) \right] \\
&= \mathbb{E}_i \left[\frac{1}{W_i^2} \left(\frac{q}{q-1} \sum_j w_{ij}^2 - \frac{1}{q-1} (\sum_j w_{ij})^2 \right) \right] \\
&= \frac{q}{q-1} \cdot \mathbb{E}_i \left[\frac{1}{W_i^2} \left(W_i^{(2)} - \frac{1}{q} W_i^2 \right) \right] \\
&= \frac{q}{q-1} \cdot \mathbb{E}_i \left[\frac{W_i^{(2)}}{W_i^2} - \frac{1}{q} \right]
\end{aligned}
$$

We then have

$$
\begin{aligned}
\left(\frac{q}{q-1} \cdot \mathbb{E}_i \left[\frac{W_i^{(2)}}{W_i^2} - \frac{1}{q} \right] \right)^{1/2} \cdot \langle u, \chi_q(h) \rangle &= \mathbb{E}_i \left[\frac{1}{W_i} \left(\sum_j w_{i,j} \langle \chi_q(j), \chi_q(h_i) \rangle \right) \right] \\
&= \mathbb{E}_i \left[\frac{1}{W_i} \left(w_{i,h_i} - \frac{\sum_{j \neq f_i} w_{i,j}}{q-1} \right) \right] \\
&= \mathbb{E}_i \left[\frac{q}{q-1} \frac{w_{i,h_i}}{W_i} - \frac{1}{q-1} \right] = \frac{q}{q-1} \cdot \mathbb{E}_i \left[\frac{w_{i,h_i}}{W_i} - \frac{1}{q} \right]
\end{aligned}
$$

For a codeword h satisfying $\mathbb{E}_i [\frac{w_{i,h_i}}{W_i}] > \frac{1}{q} + \sqrt{(1 - \frac{1}{q}) \cdot \mathbb{E}_i [\frac{W_i^{(2)}}{W_i^2} - \frac{1}{q}] \cdot \beta}$, we have $\langle u, \chi(h) \rangle > \sqrt{\beta}$ and the rest of the proof follows as in Theorem 4.4, by using the fact that any two codewords h_1, h_2 satisfy $\langle \chi(h_1), \chi(h_2) \rangle \leq \beta$. ∎

4.3 Algorithmic covering lemma

As mentioned earlier, the above lemma guarantees the existence of a distribution over the list that agrees simultaneously with the entire list, but finding such a distribution may be hard without knowing the list already! The other option is to look for a distribution over all codewords, but such a polytope will have exponential number of vertices and it's not clear how to optimize over it efficiently.

We instead relax to allow degree-t pseudodistributions, represented as degree-t pseudoexpectation operators, which can be optimized over in time $n^{O(t)}$.

We recall that $\chi : [q] \to \mathbb{R}^{q-1}$ is a function that is extended pointwise to $\chi : [q]^E \to (\mathbb{R}^{q-1})^{|E|}$. Note that χ is a 1-local vector-valued function on $[q]^E$, and so we can use $\widetilde{\mathbb{E}}[\chi(\mathbf{Z})] \in \mathbb{R}^E$ which will satisfy

$$\widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right](e) = \widetilde{\mathbb{E}}\left[\chi_q(\mathbf{Z}_e)\right]$$

4.3.1 Tanner code

Let $q \geq 2$ be an integer. Fix $G(L, R, E)$ to be an (n, d, λ)-expander, and $\mathcal{C}_0 \subseteq [q]^d$ to be an inner code. Let $\mathcal{C}^{Tan} \subseteq [q]^E$ be the Tanner code determined by G and $\mathcal{C}_0$.

Lemma 4.6. *Fix $\gamma > 0$. Let $u \in \mathbb{R}^{(q-1)|E|}$ with $\|u\|_2 = 1$. For any $t \geq d$, there exists a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ of SoS-degree t such that for any $h \in \mathcal{C}^{Tan}$ that satisfies $\langle u, \chi(h) \rangle > \gamma$,*

$$\left\langle \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right], \chi(h) \right\rangle = \mathbb{E}_{e \in E}\left[\left\langle \widetilde{\mathbb{E}}\left[\chi_q(\mathbf{Z}_e)\right], \chi_q(h(e)) \right\rangle\right] > \gamma^2$$

Moreover, this pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ can be found in time $n^{O(t)}$.

Proof. Define the convex function $\Psi(\widetilde{\mathbb{E}}[\cdot]) = \left\langle \widetilde{\mathbb{E}}[\chi(\mathbf{Z})], \widetilde{\mathbb{E}}[\chi(\mathbf{Z})] \right\rangle$ and solve the convex program in Table 1.

$$
\begin{array}{ll}
\text{minimize} & \Psi\left(\widetilde{\mathbb{E}}\left[\cdot\right]\right) \\[2mm]
\text{subject to} & \\[1mm]
& \left\langle \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right], u \right\rangle > \gamma \\[2mm]
& \widetilde{\mathbb{E}}\left[\cdot\right] \text{ is a pseudocodeword of SoS-degree } t
\end{array}
$$

Table 1: Finding cover for the list $\mathcal{L}$.

Let $\widetilde{\mathbb{E}}[\cdot]$ be the pseudocodeword obtained, with $\Psi^* = \Psi(\widetilde{\mathbb{E}}[\cdot])$. We will use the optimality of $\widetilde{\mathbb{E}}[\cdot]$ to argue that for any $h \in \mathcal{L}$,

$$\left\langle \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right], \chi(h) \right\rangle > \gamma^2$$

Suppose not. Then there exists $h \in \mathcal{C}^{Tan}$ such that $\left\langle \widetilde{\mathbb{E}}[\chi(\mathbf{Z})], \chi(h) \right\rangle \leq \gamma$. Then, for some $\theta \in [0, 1]$ to be chosen later, consider

$$\widetilde{\mathbb{E}}_\theta[\cdot] = (1 - \theta)\widetilde{\mathbb{E}}\left[\cdot\right] + \theta\,\widetilde{\mathbb{E}}^{(h)}[\cdot]$$

Here, we think of h as an SoS-degree-t pseudocodeword, so that by convexity, $\widetilde{\mathbb{E}}_\theta[\cdot]$ is also an SoS-degree-t pseudocodeword. We will show towards a contradiction that $\Psi(\widetilde{\mathbb{E}}_\theta[\cdot]) < \Psi^*$.

$$
\begin{aligned}
\Psi(\widetilde{\mathbb{E}}_\theta[\cdot]) &= \left\langle \widetilde{\mathbb{E}}_\theta[\chi(\mathbf{Z})], \widetilde{\mathbb{E}}_\theta[\chi(\mathbf{Z})] \right\rangle \\
&= \left\langle (1 - \theta) \cdot \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right] + \theta \cdot \widetilde{\mathbb{E}}^{(h)}[\chi(\mathbf{Z})], (1 - \theta) \cdot \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right] + \theta \cdot \widetilde{\mathbb{E}}^{(h)}[\chi(\mathbf{Z})] \right\rangle \\
&= \left\langle (1 - \theta) \cdot \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right] + \theta \cdot \chi(h), (1 - \theta) \cdot \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right] + \theta \cdot \chi(h) \right\rangle \\
&= (1 - \theta)^2 \cdot \left\langle \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right], \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right] \right\rangle + 2 \cdot \theta(1 - \theta) \cdot \left\langle \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right], \chi(h) \right\rangle + \theta^2 \cdot \langle \chi(h), \chi(h) \rangle \\
&\leq (1 - \theta)^2 \cdot \Psi^* + 2 \cdot \theta(1 - \theta) \cdot \gamma + \theta^2
\end{aligned}
$$

Optimizing over θ, we choose $\theta^* = \frac{\Psi^* - \gamma}{\Psi^* - 2\gamma + 1}$. As long as $\Psi^* > \gamma$, we get optimal $\theta^* > 0$, which implies $\Psi(\widetilde{\mathbb{E}}_{\theta^*}[\cdot]) < \Psi\left(\widetilde{\mathbb{E}}_\theta[\cdot]\big|_{\theta=0}\right) = \Psi^*$, which would be a contradiction. ∎

16

Lemma 4.7. *Fix $\varepsilon > 0$. Let the distance of $\mathcal{C}^{Tan}$ be δ, and let g be a received word. For any $t \geq d$, there exists a degree-t pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ such that for any $h \in \mathcal{C}^{Tan}$ such that $\Delta(g, h) < \mathcal{J}(\delta) - \varepsilon$,*

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) < \delta - 2\varepsilon \cdot \sqrt{1 - \frac{q}{q-1} \cdot \delta}$$

Moreover, this pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ can be found in time $n^{O(t)}$.

Proof. We use g to construct a vector u, which we can use to find the required pseudocodeword via Lemma 4.6.

Let $\delta = \left(1 - \frac{1}{q}\right)(1 - \beta)$ so that $\mathcal{J}(\delta) = \left(1 - \frac{1}{q}\right)(1 - \sqrt{\beta})$. Let $u = \chi(g)$.

For any $h \in \mathcal{L}(g, \mathcal{J}(\delta) - \varepsilon)$,

$$\Delta(g, h) < \mathcal{J}(\delta) - \varepsilon = \left(1 - \frac{1}{q}\right)(1 - \sqrt{\beta}) - \varepsilon \Rightarrow \langle \chi(g), \chi(h) \rangle > \sqrt{\beta} + \frac{q}{q-1} \cdot \varepsilon$$

Therefore, using Lemma 4.6 and vector u, we can find a degree-t pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ such that for any $h \in \mathcal{L}(g, \mathcal{J}(\delta) - \varepsilon)$,

$$\left\langle \widetilde{\mathbb{E}}[\chi(\mathbf{Z})], \chi(h) \right\rangle > \left(\sqrt{\beta} + \frac{q}{q-1} \cdot \varepsilon\right)^2 > \beta + 2\sqrt{\beta} \cdot \frac{q}{q-1} \cdot \varepsilon$$

Writing again in terms of distances and using $\beta = 1 - \frac{q}{q-1}\delta$, this means,

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) < \left(1 - \frac{1}{q}\right)(1 - \beta - 2\sqrt{\beta} \cdot \frac{q}{q-1} \cdot \varepsilon) = \delta - 2\sqrt{1 - \frac{q}{q-1}\delta} \cdot \varepsilon. \qquad \blacksquare$$

Next we show that Lemma 4.6 can also be used to efficiently find a cover for the list when dealing with list recovery, by using the given weights to construct a modified vector u (which will be the same vector embedding that was used in proof of Theorem 4.5).

Lemma 4.8. *Fix $\varepsilon > 0$. Let the distance of $\mathcal{C}^{Tan}$ be δ, and let the given weights be $\{w_{e,j}\}_{e \in E, j \in [q]}$. Assume that the weights are normalized so that $\sum_j w_{e,j} = 1$ and denote $W_e^{(2)} = \sum_j w_{e,j}^2$.*

For any $t \geq d$, there exists a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ of SoS-degree t such that for any $h \in \mathcal{C}^{Tan}$ that satisfies

$$\mathbb{E}_e\left[w_{e,h(e)}\right] > \frac{1}{q} + \sqrt{\left(1 - \frac{1}{q} - \delta\right) \cdot \left(\mathbb{E}_e\left[W_e^{(2)} - \frac{1}{q}\right]\right)} + \varepsilon$$

also satisfies $\Delta(\widetilde{\mathbb{E}}[\cdot], h) < \delta - \Omega_{q,\delta,W}(\varepsilon)$.

Moreover, this pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ can be found in time $n^{O(t)}$.

Proof. Again, let $\delta = \left(1 - \frac{1}{q}\right)(1 - \beta)$. We use the same vector u to embed the given weights, with the property that for any h that satisfies

$$\mathbb{E}_e\left[w_{e,h(e)}\right] > \frac{1}{q} + \sqrt{\left(1 - \frac{1}{q} - \delta\right) \cdot \left(\mathbb{E}_e\left[W_e^{(2)} - \frac{1}{q}\right]\right)} + \varepsilon$$

$$= \frac{1}{q} + \sqrt{\left(1 - \frac{1}{q}\right) \cdot \left(\mathbb{E}_e\left[W_e^{(2)} - \frac{1}{q}\right]\right) \beta} + \varepsilon$$

now satisfies

$$\langle u, \chi(h) \rangle > \sqrt{\beta} + \sqrt{\frac{q}{q-1}} \frac{\varepsilon}{\sqrt{\mathbb{E}_e[W_e^{(2)} - \frac{1}{q}]}}$$

17

Next, we appeal to Lemma 4.6 with u to efficiently obtain a degree-t pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ such that

$$\left\langle \widetilde{\mathbb{E}}\left[\chi(\mathbf{Z})\right], \chi(h) \right\rangle > \beta + 2\sqrt{\beta}\sqrt{\frac{q}{q-1}}\frac{\varepsilon}{\sqrt{\mathbb{E}_e[W_e^{(2)} - \frac{1}{q}]}}$$

In terms of distance, this means,

$$\Delta(\widetilde{\mathbb{E}}\left[\cdot\right], h) < \left(1 - \frac{1}{q}\right)\left(1 - \beta - 2\sqrt{\beta}\sqrt{\frac{q}{q-1}}\frac{\varepsilon}{\sqrt{\mathbb{E}_e[W_e^{(2)} - \frac{1}{q}]}}\right)$$

$$= \delta - 2\varepsilon \cdot \sqrt{1 - \frac{q}{q-1}\delta} \cdot \sqrt{\frac{1 - \frac{1}{q}}{\mathbb{E}_e[W_e^{(2)} - \frac{1}{q}]}}. \qquad \blacksquare$$

4.3.2 AEL Code

Fix $q \in \mathbb{N}$. Fix $G = (L, R, E)$ to be an (n, d, λ)-expander, $\mathcal{C}_0 \subseteq [q]^d$ to be an inner code, and $\mathcal{C}_1 \subseteq [|\mathcal{C}_0|]^n$ to be an outer code. Let $\mathcal{C}^{AEL} \subseteq [q^d]^R$ be the Tanner code determined by G, $\mathcal{C}_0$ and $\mathcal{C}_1$.

Definition 4.9. Let $\overline{\chi} : [q]^E \to \left(\mathbb{R}^{q^d - 1}\right)^R$ be defined as

$$(\overline{\chi}(f))(r) = \chi_{q^d}(f_{N_R(r)})$$

so that $\overline{\chi}$ is a d-local vector-valued function on $[q]^E$.

Observation 4.10. If $f_1, f_2 \in [q]^E$ with $\Delta_{AEL}(f_1, f_2) = (1 - \frac{1}{q^d})(1 - \beta)$, then $\langle \overline{\chi}(f_1), \overline{\chi}(f_2)\rangle = \beta$.

Lemma 4.11. Fix $\gamma > 0$. Let $u \in (\mathbb{R}^{q^d - 1})^R$ with $\|u\| = 1$. For any $t \geq d$, there exists a degree-t pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ such that for any $h \in \mathcal{C}^{AEL}$ that satisfies $\langle u, \overline{\chi}(h)\rangle > \gamma$,

$$\left\langle \widetilde{\mathbb{E}}\left[\overline{\chi}(\mathbf{Z})\right], \overline{\chi}(h) \right\rangle = \mathbb{E}_{r \in R}\left[\left\langle \widetilde{\mathbb{E}}\left[\chi_{q^d}(\mathbf{Z}_{N_R(r)})\right], \chi_{q^d}(h(r)) \right\rangle\right] > \gamma^2$$

Moreover, this pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ can be found in time $n^{O(t)}$.

Proof. The proof is very similar to the proof of Lemma 4.6, except we replace embedding function χ by $\overline{\chi}$.

Define the quantity $\Psi(\widetilde{\mathbb{E}}[\cdot]) = \left\langle \widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})], \widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})] \right\rangle$ and solve the following convex program.

> minimize $\quad \Psi\left(\widetilde{\mathbb{E}}\left[\cdot\right]\right)$
>
> subject to
>
> $\left\langle \widetilde{\mathbb{E}}\left[\overline{\chi}(\mathbf{Z})\right], \overline{\chi}(g) \right\rangle > \gamma$
>
> $\widetilde{\mathbb{E}}\left[\cdot\right]$ is a pseudocodeword of SoS-degree t

Table 2: Finding cover for the list $\mathcal{L}$.

Let $\widetilde{\mathbb{E}}[\cdot]$ be the pseudocodeword obtained, with $\Psi^* = \Psi(\widetilde{\mathbb{E}}[\cdot])$. Suppose it does not have the covering property. Then there exists $h \in \mathcal{L}$ such that $\left\langle \widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})], \overline{\chi}(h) \right\rangle \leq \gamma$. Then, for some $\theta \in [0, 1]$ to be chosen later, consider

$$\widetilde{\mathbb{E}}_\theta[\cdot] = (1 - \theta)\widetilde{\mathbb{E}}\left[\cdot\right] + \theta\,\widetilde{\mathbb{E}}^{(h)}[\cdot]$$

18

We again have,

$$\Psi(\widetilde{\mathbb{E}}_\theta[\cdot]) = \left\langle \widetilde{\mathbb{E}}_\theta[\overline{\chi}(\mathbf{Z})], \widetilde{\mathbb{E}}_\theta[\overline{\chi}(\mathbf{Z})] \right\rangle$$

$$= \left\langle (1-\theta)\widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})] + \theta\,\widetilde{\mathbb{E}}^{(h)}[\overline{\chi}(\mathbf{Z})], (1-\theta)\widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})] + \theta\,\widetilde{\mathbb{E}}^{(h)}[\overline{\chi}(\mathbf{Z})] \right\rangle$$

$$= \left\langle (1-\theta)\widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})] + x\overline{\chi}(h), (1-\theta)\widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})] + \theta\overline{\chi}(h) \right\rangle$$

$$= (1-\theta)^2 \left\langle \widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})], \widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})] \right\rangle + 2\theta(1-\theta) \left\langle \widetilde{\mathbb{E}}[\overline{\chi}(\mathbf{Z})], \overline{\chi}(h) \right\rangle + \theta^2 \left\langle \overline{\chi}(h), \overline{\chi}(h) \right\rangle$$

$$\leq (1-\theta)^2\Psi^* + 2\theta(1-\theta)\gamma + \theta^2$$

Optimizing over θ, we choose $\theta^* = \frac{\Psi^*-\gamma}{\Psi^*-2\gamma+1}$. We will obtain a contradiction since $\Psi^* > \gamma^2$. ∎

Lemma 4.12. *Fix $\varepsilon > 0$. Let the distance of $\mathcal{C}^{AEL}$ be δ, and let g be a received word. For any $t \geq d$, there exists a degree-t pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ such that for any $h \in \mathcal{C}^{AEL}$ such that $\Delta(g,h) < \mathcal{J}(\delta) - \varepsilon$,*

$$\mathbb{E}_r \left[\mathbf{1}\{\widetilde{\mathbb{E}}\left[\mathbf{Z}_{N_R(r)} \neq h_r\right]\} \right] < \delta - 2\varepsilon \cdot \sqrt{1 - \frac{q^d}{q^d-1} \cdot \delta}$$

Proof. Same proof as the proof of Lemma 4.7, with the alphabet changed. The received word g can be used to construct a unit vector $u = \overline{\chi}(g)$, which is then used via Lemma 4.11 to find a pseudocodeword with the required covering property. ∎

Lemma 4.13. *Fix $\varepsilon > 0$. Let the distance of $\mathcal{C}^{AEL}$ be δ, and let the given weights be $\{w_{r,j}\}_{r \in R, j \in [q^d]}$. Assume that the weights are normalized so that $\sum_j w_{r,j} = 1$ and denote $W_r^{(2)} = \sum_j w_{r,j}^2$.*

For any $t \geq d$, there exists a degree-t pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ such that for any $h \in \mathcal{C}^{AEL}$ that satisfies

$$\mathbb{E}_r \left[w_{r,h(r)} \right] > \frac{1}{q^d} + \sqrt{\left(1 - \frac{1}{q^d} - \delta\right) \cdot \left(\mathbb{E}_r\left[W_r^{(2)} - \frac{1}{q^d}\right]\right)} + \varepsilon$$

also satisfies $\Delta(\widetilde{\mathbb{E}}[\cdot], h) < \delta - \Omega_{q,d,\delta,W}(\varepsilon)$.

Moreover, this pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ can be found in time $n^{O(t)}$.

Proof. Same proof as the proof of Lemma 4.8. ∎

5 Sum-of-Squares Proofs of Distance

We will be proving that pseudocodewords satisfying certain η-good property defined below have the same distance properties as true codewords, up to η error.

Definition 5.1. *A pseudocodeword of SoS-degree at least $2d$ is η-good if*

$$\mathbb{E}_{\ell,r} \left[\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)}] \right] \leq \eta$$

Observation 5.2. *A true codeword is a 0-good pseudocodeword.*

The η-good property is useful to change the pseudoexpectation of a product of functions into the product of pseudoexpectations of those functions. We establish a formal claim about this in the next lemma. The terms $\|X_\ell\|_\infty$ and $\|Y_r\|_\infty$ should be just seen as normalizing the scale, and indeed we will only use functions that are bounded in infinity norm by 1.

19

Lemma 5.3. *Let $\{X_\ell\}_{\ell \in L}$ and $\{Y_r\}_{r \in R}$ be two collections of d-local functions on $[q]^E$ such that for every $\ell \in L$, $X_\ell(f)$ only depends on $f|_{N_L(\ell)}$ and for every $r \in R$, $Y_r(f)$ only depends on $f|_{N_R(r)}$. Then, for an η-good pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$,*

$$\mathop{\mathbb{E}}_{\ell,r}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})Y_r(\mathbf{Z})\right]\right] \;\leq\; \mathop{\mathbb{E}}_{\ell,r}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})\right]\widetilde{\mathbb{E}}\left[Y_r(\mathbf{Z})\right]\right] + \eta\left(\max_\ell \|X_\ell\|_\infty\right)\left(\max_r \|Y_r\|_\infty\right)$$

Proof. For any ℓ and r,

$$\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})Y_r(\mathbf{Z})\right] - \widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})\right]\widetilde{\mathbb{E}}\left[Y_r(\mathbf{Z})\right]$$

$$= \sum_{\substack{\alpha \in [q]^{N_L(\ell)} \\ \beta \in [q]^{N_R(r)}}} \widetilde{\mathbb{E}}\left[X_\ell(\alpha)\prod_{s \in N_L(\ell)}\mathbf{Z}_{s,\alpha_s}\cdot Y_r(\beta)\prod_{t \in N_R(r)}\mathbf{Z}_{t,\beta_t}\right] - \sum_{\substack{\alpha \in [q]^{N_L(\ell)} \\ \beta \in [q]^{N_R(r)}}}\cdot\widetilde{\mathbb{E}}\left[X_\ell(\alpha)\prod_{s \in N_L(\ell)}\mathbf{Z}_{s,\alpha_s}\right]\widetilde{\mathbb{E}}\left[Y_r(\beta)\prod_{t \in N_R(r)}\mathbf{Z}_{t,\beta_t}\right]$$

$$= \sum_{\alpha,\beta}X_\ell(\alpha)Y_r(\beta)\cdot\left(\widetilde{\mathbb{E}}\left[\prod_{s \in N_L(\ell)}\mathbf{Z}_{s,\alpha_s}\cdot\prod_{t \in N_R(r)}\mathbf{Z}_{t,\beta_t}\right] - \widetilde{\mathbb{E}}\left[\prod_{s \in N_L(\ell)}\mathbf{Z}_{s,\alpha_s}\right]\cdot\widetilde{\mathbb{E}}\left[\prod_{t \in N_R(r)}\mathbf{Z}_{t,\beta_t}\right]\right)$$

$$\leq \|X_\ell\|_\infty\|Y_r\|_\infty\cdot\sum_{\alpha,\beta}\left|\widetilde{\mathbb{E}}\left[\prod_{s \in N_L(\ell)}\mathbf{Z}_{s,\alpha_s}\prod_{t \in N_R(r)}\mathbf{Z}_{t,\beta_t}\right] - \widetilde{\mathbb{E}}\left[\prod_{s \in N_L(\ell)}\mathbf{Z}_{s,\alpha_s}\right]\cdot\widetilde{\mathbb{E}}\left[\prod_{t \in N_R(r)}\mathbf{Z}_{t,\beta_t}\right]\right|$$

$$= \|X_\ell\|_\infty\|Y_r\|_\infty\cdot\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell)},\mathbf{Z}_{N_R(r)}]$$

Averaging over ℓ and r, we get

$$\mathop{\mathbb{E}}_{\ell,r}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})Y_r(\mathbf{Z})\right] - \widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})\right]\widetilde{\mathbb{E}}\left[Y_r(\mathbf{Z})\right]\right] \leq \mathop{\mathbb{E}}_{\ell,r}\left[\|X_\ell\|_\infty\|Y_r\|_\infty\cdot\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell)},\mathbf{Z}_{N_R(r)}]\right]$$

$$\leq \max_\ell\|X_\ell\|_\infty\cdot\max_r\|Y_r\|_\infty\cdot\mathop{\mathbb{E}}_{\ell,r}\left[\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell)},\mathbf{Z}_{N_R(r)}]\right]$$

$$\leq \eta\cdot\max_\ell\|X_\ell\|_\infty\cdot\max_r\|Y_r\|_\infty \qquad\blacksquare$$

The proofs of distance for both Tanner and AEL codes go via Expander Mixing Lemma (EML), and so we establish the analog of EML for pseudocodewords. Morally speaking, EML allows us to change measure from the edges of an expander to the complete (bipartite) graph for product functions. First we prove a version of EML for vector valued functions, and then show that since pseudoexpectation operators can be written in terms of certain underlying vectors, they also satisfy a version of EML. Note that this step does not require any η-good property.

Lemma 5.4 (EML for vector-valued functions). *Let $\{v_\ell\}_{\ell \in L}$ and $\{u_r\}_{r \in R}$ be a collection of vectors in $\mathbb{R}^N$. Then,*

$$\left|\mathop{\mathbb{E}}_{\ell \sim r}\left[\langle v_\ell, u_r\rangle\right] - \mathop{\mathbb{E}}_{\ell,r}\left[\langle v_\ell, u_r\rangle\right]\right| \leq \lambda\sqrt{\mathop{\mathbb{E}}_\ell\left[\|v_\ell\|^2\right]}\sqrt{\mathop{\mathbb{E}}_r\left[\|u_r\|^2\right]}$$

Proof. Usual EML applied coordinate-wise. $\qquad\blacksquare$

Lemma 5.5 (EML for pseudoexpectations). *Let $\{X_\ell\}_{\ell \in L}$ and $\{Y_r\}_{r \in R}$ be two collections of d-local functions on $[q]^E$ such that for every $\ell \in L$, $X_\ell(f)$ only depends on $f|_{N_L(\ell)}$ and for every $r \in R$, $Y_r(f)$ only depends on $f|_{N_R(r)}$. Then for a λ-spectral expander, we have*

$$\left|\mathop{\mathbb{E}}_{\ell \sim r}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})Y_r(\mathbf{Z})\right]\right] - \mathop{\mathbb{E}}_{\ell,r}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})Y_r(\mathbf{Z})\right]\right]\right| \leq \lambda\sqrt{\mathop{\mathbb{E}}_\ell\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})^2\right]\right]}\sqrt{\mathop{\mathbb{E}}_r\left[\widetilde{\mathbb{E}}\left[Y_r(\mathbf{Z})^2\right]\right]}.$$

Proof. Consider the $2n \times 2n$ matrix M, with

$$M_{ij} = \begin{cases} \widetilde{\mathbb{E}}[X_iX_j], & 1 \leq i \leq n, 1 \leq j \leq n \\ \widetilde{\mathbb{E}}[X_iY_{j-n}], & 1 \leq i \leq n, n+1 \leq j \leq 2n \\ \widetilde{\mathbb{E}}[Y_{i-n}X_j], & n+1 \leq i \leq 2n, 1 \leq j \leq n \\ \widetilde{\mathbb{E}}[Y_{i-n}Y_{j-n}] & n+1 \leq i \leq 2n, n+1 \leq j \leq 2n \end{cases}$$

For any vector $v = (x_1, x_2, \cdots, x_n, y_1, y_2, \cdots, y_n)$, we show that $v^T M v \geq 0$, so that M is PSD.

$$
\begin{aligned}
v^T M v &= \mathop{\mathbb{E}}_{i,j} \left[M_{ij} x_i x_j + M_{i,j+n} x_i y_j + M_{i+n,j} y_i x_j + M_{i+n,j+n} y_i y_j \right] \\
&= \mathop{\mathbb{E}}_{i,j} \left[\widetilde{\mathbb{E}} \left[X_i X_j \right] x_i x_j + \widetilde{\mathbb{E}} \left[X_i Y_j \right] x_i y_j + \widetilde{\mathbb{E}} \left[Y_i X_j \right] y_i x_j + \widetilde{\mathbb{E}} \left[Y_i Y_j \right] y_i y_j \right] \\
&= \mathop{\mathbb{E}}_{i,j} \left[\widetilde{\mathbb{E}} \left[x_i x_j X_i X_j \right] + \widetilde{\mathbb{E}} \left[x_i y_j X_i Y_j \right] + \widetilde{\mathbb{E}} \left[y_i x_j Y_i X_j \right] + \widetilde{\mathbb{E}} \left[y_i y_j Y_i Y_j \right] \right] \\
&= \mathop{\mathbb{E}}_{i,j} \left[\widetilde{\mathbb{E}} \left[(x_i X_i + y_i Y_i)(x_j X_j + y_j Y_j) \right] \right] \\
&= \widetilde{\mathbb{E}} \left[\mathop{\mathbb{E}}_{i,j} \left[(x_i X_i + y_i Y_i)(x_j X_j + y_j Y_j) \right] \right] \\
&= \widetilde{\mathbb{E}} \left[\mathop{\mathbb{E}}_{i} \left[(x_i X_i + y_i Y_i) \right]^2 \right] \geq 0
\end{aligned}
$$

Therefore there exist vectors $\{v_\ell\}_{\ell \in L}$ and $\{u_r\}_{r \in R}$ such that

$$
\widetilde{\mathbb{E}} \left[X_\ell Y_r \right] = \langle v_\ell, u_r \rangle, \quad \widetilde{\mathbb{E}} \left[X_\ell^2 \right] = \langle v_\ell, v_\ell \rangle, \quad \text{and} \quad \widetilde{\mathbb{E}} \left[Y_r^2 \right] = \langle u_r, u_r \rangle
$$

Applying Lemma 5.4 to the collection of vectors obtained above, we immediately obtain,

$$
\left| \mathop{\mathbb{E}}_{\ell \sim r} \left[\widetilde{\mathbb{E}} \left[X_\ell Y_r \right] \right] - \mathop{\mathbb{E}}_{\ell, r} \left[\widetilde{\mathbb{E}} \left[X_\ell Y_r \right] \right] \right| \leq \lambda \sqrt{\mathop{\mathbb{E}}_{\ell} \left[\widetilde{\mathbb{E}} \left[X_\ell^2 \right] \right]} \sqrt{\mathop{\mathbb{E}}_{r} \left[\widetilde{\mathbb{E}} \left[Y_r^2 \right] \right]} \qquad \blacksquare
$$

5.1 Tanner Code

Suppose we are working with a Tanner code $\mathcal{C}^{Tan}$ with inner code $\mathcal{C}_0$ of distance δ_0, so that the distance of $\mathcal{C}^{Tan}$ is at least $\delta_0(\delta_0 - \lambda)$. We show that η-good pseudocodewords satisfy a similar distance property, up to error η.

Lemma 5.6 (Distance of Tanner code). *The distance between an η-good pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ and a true codeword h is at least $\delta_0(\delta_0 - \lambda) - 2\eta \frac{\delta_0}{\delta_0 - \lambda}$, or at most $\frac{4\eta^2}{(\delta_0 - \lambda)^2} + \frac{\eta(\delta_0 + \lambda)}{\delta_0 - \lambda}$. In particular, if $\lambda \leq \delta_0/3$ and $\eta \leq \delta_0^2/9$, then $\Delta(h, \widetilde{\mathbb{E}}[\cdot]) \leq 3\eta$ or $\Delta(h, \widetilde{\mathbb{E}}[\cdot]) \geq \delta_0(\delta_0 - \lambda) - 3\eta$.*

Proof. Let $X_\ell(\mathbf{Z}) := \mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}$ and $Y_r(\mathbf{Z}) := \mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}$, and let τ denote the quantity $\sqrt{\mathbb{E}_\ell[\widetilde{\mathbb{E}}[X_\ell(\mathbf{Z})]]} \cdot \sqrt{\mathbb{E}_r[\widetilde{\mathbb{E}}[Y_r(\mathbf{Z})]]}$. Then, we have

$$
\begin{aligned}
\Delta \left(h, \widetilde{\mathbb{E}}[\cdot] \right) &= \mathop{\mathbb{E}}_{e} \left[\widetilde{\mathbb{E}} \left[\mathbf{1}\{\mathbf{Z}_e \neq h_e\} \right] \right] \\
&\leq \mathop{\mathbb{E}}_{\ell \sim r} \left[\widetilde{\mathbb{E}} \left[X_\ell(\mathbf{Z}) \cdot Y_r(\mathbf{Z}) \right] \right] \\
&\leq \mathop{\mathbb{E}}_{\ell, r} \left[\widetilde{\mathbb{E}} \left[X_\ell(\mathbf{Z}) \cdot Y_r(\mathbf{Z}) \right] \right] + \lambda \cdot \sqrt{\mathop{\mathbb{E}}_{\ell} \left[\widetilde{\mathbb{E}} \left[X_\ell(\mathbf{Z})^2 \right] \right]} \cdot \sqrt{\mathop{\mathbb{E}}_{r} \left[\widetilde{\mathbb{E}} \left[Y_r(\mathbf{Z})^2 \right] \right]} \\
&= \mathop{\mathbb{E}}_{\ell, r} \left[\widetilde{\mathbb{E}} \left[X_\ell(\mathbf{Z}) \cdot Y_r(\mathbf{Z}) \right] \right] + \lambda \cdot \sqrt{\mathop{\mathbb{E}}_{\ell} \left[\widetilde{\mathbb{E}} \left[X_\ell(\mathbf{Z}) \right] \right]} \cdot \sqrt{\mathop{\mathbb{E}}_{r} \left[\widetilde{\mathbb{E}} \left[Y_r(\mathbf{Z}) \right] \right]} \\
&= \mathop{\mathbb{E}}_{\ell, r} \left[\widetilde{\mathbb{E}} \left[X_\ell(\mathbf{Z}) \cdot Y_r(\mathbf{Z}) \right] \right] + \lambda \cdot \tau \\
&= \mathop{\mathbb{E}}_{\ell, r} \left[\widetilde{\mathbb{E}} \left[X_\ell(\mathbf{Z}) \right] \cdot \widetilde{\mathbb{E}} \left[Y_r(\mathbf{Z}) \right] \right] + \eta \cdot 1 \cdot 1 + \lambda \cdot \tau \\
&= \tau^2 + \lambda \cdot \tau + \eta
\end{aligned}
$$

21

On the other hand,

$$
\begin{aligned}
\Delta\left(h, \widetilde{\mathbb{E}}\left[\cdot\right]\right) &= \underset{e}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_e \neq h_e\}\right]\right] \\
&= \underset{\ell}{\mathbb{E}}\left[\underset{e \in N_L(\ell)}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_e \neq h_e\}\right]\right]\right] \\
&= \underset{\ell}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[\Delta\left(\mathbf{Z}_{N_L(\ell)}, h_{N_L(\ell)}\right)\right]\right] \\
&\geq \underset{\ell}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[0 \cdot \mathbf{1}\{\mathbf{Z}_{N_L(\ell)} = h_{N_L(\ell)}\} + \delta_0 \cdot \mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\right]\right] \\
&= \underset{\ell}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[\delta_0 \cdot \mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\right]\right] \\
&= \underset{\ell}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[\delta_0 \cdot X_\ell(\mathbf{Z})\right]\right] \\
&= \delta_0 \cdot \underset{\ell}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})\right]\right]
\end{aligned}
$$

Likewise, $\Delta\left(h, \widetilde{\mathbb{E}}[\cdot]\right) \geq \delta_0 \cdot \mathbb{E}_r[\widetilde{\mathbb{E}}[Y_r(\mathbf{Z})]]$, and so,

$$
\Delta\left(h, \widetilde{\mathbb{E}}\left[\cdot\right]\right) \geq \delta_0 \cdot \sqrt{\underset{\ell}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[X_\ell(\mathbf{Z})\right]\right]} \cdot \sqrt{\underset{r}{\mathbb{E}}\left[\widetilde{\mathbb{E}}\left[Y_r(\mathbf{Z})\right]\right]} = \delta_0 \cdot \tau
$$

Comparing, we get, $\tau^2 + \lambda \cdot \tau + \eta \geq \delta_0 \cdot \tau$, which means,

$$
\tau \geq \frac{(\delta_0 - \lambda) + \sqrt{(\delta_0 - \lambda)^2 - 4\eta}}{2} \qquad \text{or} \qquad \tau \leq \frac{(\delta_0 - \lambda) - \sqrt{(\delta_0 - \lambda)^2 - 4\eta}}{2}
$$

In the first case, we have

$$
\begin{aligned}
\Delta\left(h, \widetilde{\mathbb{E}}\left[\cdot\right]\right) &\geq \delta_0 \cdot \tau \geq \delta_0 \cdot \frac{(\delta_0 - \lambda) + \sqrt{(\delta_0 - \lambda)^2 - 4\eta}}{2} \\
&= \frac{\delta_0(\delta_0 - \lambda)}{2}\left(1 + \sqrt{1 - \frac{4\eta}{(\delta_0 - \lambda)^2}}\right) \\
&\geq \frac{\delta_0(\delta_0 - \lambda)}{2}\left(1 + 1 - \frac{4\eta}{(\delta_0 - \lambda)^2}\right) \\
&= \delta_0(\delta_0 - \lambda) - 2\eta \cdot \frac{\delta_0}{\delta_0 - \lambda} \ .
\end{aligned}
$$

Also, in the second case, we have

$$
\begin{aligned}
\tau &\leq \frac{(\delta_0 - \lambda) - \sqrt{(\delta_0 - \lambda)^2 - 4\eta}}{2} = \frac{\delta_0 - \lambda}{2}\left(1 - \sqrt{1 - \frac{4\eta}{(\delta_0 - \lambda)^2}}\right) \\
&\leq \frac{\delta_0 - \lambda}{2}\left(1 - 1 + \frac{4\eta}{(\delta_0 - \lambda)^2}\right) \\
&= \frac{2\eta}{\delta_0 - \lambda} \ ,
\end{aligned}
$$

which gives

$$
\Delta(g, \widetilde{\mathbb{E}}\left[\cdot\right]) \leq \tau^2 + \lambda\tau + \eta \leq \frac{4\eta^2}{(\delta_0 - \lambda)^2} + \frac{2\eta\lambda}{\delta_0 - \lambda} + \eta = \frac{4\eta^2}{(\delta_0 - \lambda)^2} + \frac{\eta(\delta_0 + \lambda)}{\delta_0 - \lambda} \qquad \blacksquare
$$

22

5.2 AEL Code

Let $\mathcal{C}_1$ be an outer code on an (n, d, λ)-expander graph $G(L, R, E)$ and let $\mathcal{C}_0$ be the inner code. Let $\mathcal{C}^{AEL}$ be the code obtained by redistributing symbols along the edges of G and then collecting them on vertices of R, as explained in Section 2.2.

Let δ_0 be the distance of $\mathcal{C}_0$, so that (designed) distance of $\mathcal{C}^{AEL}$ is $\delta = \delta_0 - \frac{\lambda}{\delta_1}$. Let $h \in [q_0]^E$ be a codeword in $\mathcal{C}^{AEL}$. We show that an η-good pseudocodeword that has some left-distance from h has a much larger right-distance from h.

Lemma 5.7 (Distance of AEL Code). *For an η-good pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ and a codeword $h \in \mathcal{C}^{AEL}$,*

$$\Delta^R(\widetilde{\mathbb{E}}[\cdot], h) \geq \delta_0 - \frac{\lambda + \eta}{\Delta^L(\widetilde{\mathbb{E}}[\cdot], h)}$$

Proof. We establish upper and lower bounds on $\Delta(\widetilde{\mathbb{E}}[\cdot], h)$.

$$
\begin{aligned}
\Delta(\widetilde{\mathbb{E}}[\cdot], h) &= \mathop{\mathbb{E}}_{e}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_e \neq h_e\}\right]\right] \\
&\geq \mathop{\mathbb{E}}_{\ell \in L}\left[\widetilde{\mathbb{E}}\left[0 \cdot \mathbf{1}\{\mathbf{Z}_{N_L(\ell)} = h_{N_L(\ell)}\} + \delta_0 \cdot \mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\right]\right] \\
&= \delta_0 \mathop{\mathbb{E}}_{\ell \in L}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\right]\right] \\
&= \delta_0 \Delta^L(\widetilde{\mathbb{E}}[\cdot], h)
\end{aligned}
$$

For the upper bound, we again rely on the expander mixing lemma:

$$
\begin{aligned}
\Delta(\widetilde{\mathbb{E}}[\cdot], h) &= \mathop{\mathbb{E}}_{e}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_e \neq h_e\}\right]\right] \\
&\leq \mathop{\mathbb{E}}_{\ell \sim r}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}\right]\right] \\
&\leq \mathop{\mathbb{E}}_{\ell, r}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}\right]\right] + \lambda \\
&\leq \mathop{\mathbb{E}}_{\ell, r}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\right]\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}\right]\right] + \lambda + \eta \\
&= \mathop{\mathbb{E}}_{\ell}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\right]\right]\mathop{\mathbb{E}}_{r}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}\right]\right] + \lambda + \eta \\
&= \Delta^L(\widetilde{\mathbb{E}}[\cdot], h) \cdot \Delta^R(\widetilde{\mathbb{E}}[\cdot], h) + \lambda + \eta
\end{aligned}
$$

Dividing the two bounds by $\Delta^L(\widetilde{\mathbb{E}}[\cdot], h)$ and rearranging, we finally get,

$$\Delta^R(\widetilde{\mathbb{E}}[\cdot], h) \geq \delta_0 - \frac{\lambda + \eta}{\Delta^L(\widetilde{\mathbb{E}}[\cdot], h)}. \qquad \blacksquare$$

6 Correlation Reduction via Conditioning

We will use the following claim from [BRS11] (see Lemma 5.2 there) that says that if $\mathbf{Z}_S$ and $\mathbf{Z}_T$ have a large covariance, then conditioning on $\mathbf{Z}_T$ reduces the variance of $\mathbf{Z}_S$ significantly.

Lemma 6.1. *Let $\widetilde{\mathbb{E}}[\cdot]$ be a pseudoexpectation operator of SoS-degree t with associated pseudocovariance and pseudovariance operators. Assume S, T are sets such that $|S| + |T| \leq t/2$, then,*

$$\widetilde{\mathrm{Var}}\left[\mathbf{Z}_S | \mathbf{Z}_T\right] \leq \widetilde{\mathrm{Var}}\left[\mathbf{Z}_S\right] - \frac{1}{q^{|T|}} \sum_{\alpha \in [q]^S, \beta \in [q]^T} \frac{(\widetilde{\mathrm{Cov}}[\mathbf{Z}_{S,\alpha}, \mathbf{Z}_{T,\beta}])^2}{\widetilde{\mathrm{Var}}[\mathbf{Z}_{T,\beta}]}$$

In particular, observe that pseudovariances are non-increasing under conditioning. The next lemma shows that if the average covariance across all pairs (ℓ, r) is η, then conditioning on a random vertex in R will reduce the average variance in L in expectation by $\Omega(\eta^2)$. Then, Lemma 6.3 will use that this cannot happen more than $\mathcal{O}(1/\eta^2)$ times, and then we must end up with a conditioned pseudoexpectation operator which has low average covariance, that is, it is η-good.

Lemma 6.2. *Let* $\eta < \mathbb{E}_{\ell,r}[\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)}]]$. *Then,*

$$\mathbb{E}_{r \in R}\left[\mathbb{E}_{\ell}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}|\mathbf{Z}_{N_R(r)}\right]\right]\right] < \mathbb{E}_{\ell}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}\right]\right] - \frac{1}{q^{2d}}\eta^2$$

Proof.

$$
\begin{aligned}
\mathbb{E}_{r \in R}\left[\mathbb{E}_{\ell}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}|\mathbf{Z}_{N_R(r)}\right]\right]\right] &= \mathbb{E}_{\ell,r}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}|\mathbf{Z}_{N_R(r)}\right]\right] \\
&\leq \mathbb{E}_{\ell,r}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}\right] - \frac{1}{q^d}\sum_{\alpha,\beta}\frac{(\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell),\alpha}, \mathbf{Z}_{N_R(r),\beta}])^2}{\widetilde{\mathrm{Var}}[\mathbf{Z}_{N_R(r),\beta}]}\right] \\
&\leq \mathbb{E}_{\ell,r}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}\right] - \frac{1}{q^d}\sum_{\alpha,\beta}\left(\widetilde{\mathrm{Cov}}\left[\mathbf{Z}_{N_L(\ell),\alpha}, \mathbf{Z}_{N_R(r),\beta}\right]\right)^2\right] \\
&\leq \mathbb{E}_{\ell,r}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}\right] - \frac{1}{q^{3d}}\left(\sum_{\alpha,\beta}\left|\widetilde{\mathrm{Cov}}\left[\mathbf{Z}_{N_L(\ell),\alpha}, \mathbf{Z}_{N_R(r),\beta}\right]\right|\right)^2\right] \\
&= \mathbb{E}_{\ell}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}\right]\right] - \frac{1}{q^{3d}}\mathbb{E}_{\ell,r}\left[\left(\widetilde{\mathrm{Cov}}\left[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)}\right]\right)^2\right] \\
&\leq \mathbb{E}_{\ell}\left[\widetilde{\mathrm{Var}}\left[\mathbf{Z}_{N_L(\ell)}\right]\right] - \frac{1}{q^{3d}}\left(\mathbb{E}_{\ell,r}\left[\widetilde{\mathrm{Cov}}\left[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)}\right]\right]\right)^2 \quad \blacksquare
\end{aligned}
$$

Lemma 6.3. *Let* $\eta > 0$ *be arbitrarily small. Given any SoS solution* $\widetilde{\mathbb{E}}[\cdot]$ *of degree* $\geq 2d\left(\frac{q^{3d}}{\eta^2} + 1\right)$, *there exists a number* $k^* \leq q^{3d}/\eta^2$ *such that*

$$\mathbb{E}_{v_1,v_2,\cdots,v_{k^*}}\left[\mathbb{E}_{\ell,r}\left[\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)}|\mathbf{Z}_{N_R(v_1)}, \mathbf{Z}_{N_R(v_2)}, \cdots, \mathbf{Z}_{N_R(v_{k^*})}]\right]\right] \leq \eta$$

Proof. Consider $\Phi_k(\widetilde{\mathbb{E}}[\cdot]) = \mathbb{E}_{v_1,v_2,\cdots,v_k}[\mathbb{E}_{\ell}[\widetilde{\mathrm{Var}}[\mathbf{Z}_{N_L(\ell)}|\mathbf{Z}_{N_R(v_1)}, \mathbf{Z}_{N_R(v_2)}, \cdots, \mathbf{Z}_{N_R(v_k)}]]]$. We know that

$$1 \geq \Phi_0 \geq \Phi_1 \geq \cdots \geq \Phi_{q^{3d}/\eta^2} \geq 0$$

so there exists a $k^* \leq q^{3d}/\eta^2$ such that $\Phi_{k^*}(\widetilde{\mathbb{E}}[\cdot]) - \Phi_{k^*+1}(\widetilde{\mathbb{E}}[\cdot]) \leq \frac{1}{q^{3d}/\eta^2} = \eta^2/q^{3d}$. By contrapositive of Lemma 6.2, this means that

$$\mathbb{E}_{v_1,v_2,\cdots,v_{k^*}}\left[\mathbb{E}_{\ell,r}\left[\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)}|\mathbf{Z}_{N_R(v_1)}, \mathbf{Z}_{N_R(v_2)}, \cdots, \mathbf{Z}_{N_R(v_{k^*})}]\right]\right] \leq \eta \quad \blacksquare$$

7 List Decoding up to Johnson Bound

In this section, we combine different pieces of the proof to give list decoding algorithms up to the Johnson bound. Note that in both Tanner and AEL cases, we reduce to unique decoding of either the same code or the base code, and this unique decoding needs to be done from pseudocodewords instead of codewords. We will handle this slight strengthening of unique decoding via randomized rounding in Section 7.3.

7.1 Tanner code

Let $\mathcal{C}^{Tan}$ be a Tanner code on an (n, d, λ)-expander graph $G(L, R, E)$, with $\mathcal{C}_0$ as the inner code. Let δ_0 be the distance of $\mathcal{C}_0$, so that (designed) distance of $\mathcal{C}^{Tan}$ is $\delta = \delta_0(\delta_0 - \lambda)$. Assume $\lambda \leq \delta_0/3$. Given $g \in [q]^E$, we wish to recover the list $\mathcal{L}(g, \mathcal{J}(\delta) - \varepsilon)$. As $\varepsilon \to 0$, the decoding radius gets arbitrarily close to the Johnson bound.

Theorem 7.1 (List decoding Tanner codes). *There is a deterministic algorithm based on $\mathcal{O}_{q,d}(1/\varepsilon^4)$ levels of the SoS-hierarchy that given g runs in time $n^{\mathcal{O}_{q,d}(1/\varepsilon^4)}$ time and computes the list $\mathcal{L}(g, \mathcal{J}(\delta) - \varepsilon)$.*

Proof. We apply the algorithmic covering Lemma 4.7 to obtain a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ of SoS-degree $t \geq 2d \left(\frac{q^{3d}}{\eta^2} + 2 \right)$ such that for any $h \in \mathcal{L}(g, \mathcal{J}(\delta) - \varepsilon)$, we know that $\Delta(\widetilde{\mathbb{E}}[\cdot], h) \leq \delta - \varepsilon_2$ for $\varepsilon_2 = 2\varepsilon \cdot \sqrt{1 - \frac{q}{q-1}\delta} \geq \Omega(\varepsilon)$. We will choose η later, and note that the choice of η does not change ε_2. Henceforth, we fix an $h \in \mathcal{L}((g, \mathcal{J}(\delta) - \varepsilon)$, so that $\Delta(\widetilde{\mathbb{E}}[\cdot], h) \leq \delta - \varepsilon_2$. Our goal is to recover h.

Pick a random $k \in \{1, \cdots, \lceil \frac{q^{3d}}{\eta^2} \rceil\}$. From Lemma 6.3, we know that with probability at least $\frac{\eta^2}{q^{3d}}$,

$$\mathbb{E}_{v_1, v_2, \cdots, v_k} \left[\mathbb{E}_{\ell, r} \left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(v_1)}, \mathbf{Z}_{N_R(v_2)}, \cdots, \mathbf{Z}_{N_R(v_k)}] \right] \right] \leq \eta . \tag{1}$$

We assume that we found a k such that Eq. (1) holds. Let V be the (random) set of k vertices we condition on, that is, $V = \{v_1, v_2, \cdots, v_k\}$, and let $N_R(V) = \cup_{v \in V} N_R(v)$. Then,

$$\mathbb{E}_{V \subseteq R, |V|=k} \left[\mathbb{E}_{\ell, r} \left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(V)}] \right] \right] \leq \eta$$

More explicitly, conditioning on $N_R(V)$ involves sampling an assignment for $N_R(V)$ according to the local distribution of $\mathbf{Z}_{N_R(V)}$. Let this random assignment be β, and we get

$$\mathbb{E}_{\substack{V \subseteq R, |V|=k \\ \beta \sim \mathbf{Z}_{N_R(V)}}} \left[\mathbb{E}_{\ell, r} \left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(V)} = \beta] \right] \right] \leq \eta$$

By Markov's inequality,

$$\mathbb{P}_{\substack{V \subseteq R, |V|=k \\ \beta \sim \mathbf{Z}_{N_R(V)}}} \left[\mathbb{E}_{\ell, r} \left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(V)} = \beta] \right] > \frac{\varepsilon_2}{15} \right] \leq \frac{15\eta}{\varepsilon_2}$$

By choosing $\eta = \varepsilon_2^2/60\delta$, we get

$$\mathbb{P}_{\substack{V \subseteq R, |V|=k \\ \beta \sim \mathbf{Z}_{N_R(V)}}} \left[\mathbb{E}_{\ell, r} \left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(V)} = \beta] \right] \leq \frac{\varepsilon_2}{15} \right] \geq 1 - \varepsilon_2/4\delta \tag{2}$$

For some V, β such that Eq. (2) holds, we will be using $\widetilde{\mathbb{E}}[\ \cdot \ | \mathbf{Z}_{N_R(V)} = \beta]$ as an η-good pseudocodeword. Using $\varepsilon_2 < \delta$, note that

$$\frac{\varepsilon_2}{15} < \frac{\delta}{15} = \frac{\delta_0 \cdot (\delta_0 - \lambda)}{15} < \frac{\delta_0^2}{9}$$

and $\lambda \leq \delta_0/3$ so that the conditions of Lemma 5.6 are satisfied.

For this η-good pseudocodeword, we need to argue that it is still close to h that we are trying to find. This is easy to ensure in expectation, and we again appeal to Markov's inequality to say that it

also holds with significant probability, up to some loss in distance. By the law of total expectation, for any $V \subseteq R$,

$$\mathbb{E}_{\beta \sim \mathbf{Z}_{N_R(V)}} \left[\Delta \left(\widetilde{\mathbb{E}} \left[\; \cdot \; | \mathbf{Z}_{N_R(V)} = \beta \right], h \right) \right] = \Delta(\widetilde{\mathbb{E}} \left[\cdot \right], h) \leq \delta - \varepsilon_2$$

Averaging over all $V \subseteq R$ of size k,

$$\mathbb{E}_{\substack{V \subseteq R, |V|=k \\ \beta \sim \mathbf{Z}_{N_R(V)}}} \left[\Delta \left(\widetilde{\mathbb{E}} \left[\; \cdot \; | \mathbf{Z}_{N_R(V)} = \beta \right], h \right) \right] \leq \delta - \varepsilon_2 \tag{3}$$

Again, we claim via Markov's inequality that a significant fraction of conditionings must end up being not too far from f.

$$\mathbb{P}_{\substack{V \subseteq R, |V|=k \\ \beta \sim \mathbf{Z}_{N_R(V)}}} \left[\Delta \left(\widetilde{\mathbb{E}} \left[\; \cdot \; | \mathbf{Z}_{N_R(V)} = \beta \right], h \right) \leq \delta - \frac{\varepsilon_2}{2} \right] \geq \frac{\varepsilon_2/2}{\delta - \varepsilon_2 + \frac{\varepsilon_2}{2}} \geq \frac{\varepsilon_2}{2\delta} \tag{4}$$

Henceforth, we fix a conditioning (V, β) with $\beta \in [q]^{N_R(V)}$ such that events in both Eq. (2) and Eq. (4) happen. Note that by a union bound, a random (V, β) has this property with probability at least $\varepsilon_2/4\delta$. Fix such a conditioning, and let the conditioned pseudoexpectation be

$$\widetilde{\mathbb{E}}^* \left[\cdot \right] = \widetilde{\mathbb{E}} \left[\; \cdot \; | \mathbf{Z}_{N_R(V)=\beta} \right]$$

and the corresponding covariance operator be $\widetilde{\mathrm{Cov}}_{\widetilde{\mathbb{E}}^*}[\cdot]$ Note that the degree of $\widetilde{\mathbb{E}}^*[\cdot]$ is $t - 2d \cdot k \geq 4d$. From definition, we know that

$$\Delta(\widetilde{\mathbb{E}}^* \left[\cdot \right], h) \leq \delta - \frac{\varepsilon_2}{2} \tag{5}$$

$$\mathbb{E}_{\ell, r} \left[\widetilde{\mathrm{Cov}}_{\widetilde{\mathbb{E}}^*}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)}] \right] \leq \frac{\varepsilon_2}{15} \tag{6}$$

From Eq. (6) and Lemma 5.6, we know that $\Delta(\widetilde{\mathbb{E}}^*[\cdot], h) \leq \varepsilon_2/5$ or $\Delta(\widetilde{\mathbb{E}}^*[\cdot], h) \geq \delta - \varepsilon_2/5$. The latter is impossible because of Eq. (5), and so we must have $\Delta(\widetilde{\mathbb{E}}^*[\cdot], h) \leq \varepsilon_2/5 < \delta/5$.

Finally, we use Lemma 7.7 to recover h using $\widetilde{\mathbb{E}}^*[\cdot]$ with probability at least $1/5$. The algorithm succeeds if

1. a k is picked so that Eq. (1) holds,

2. (V, β) is picked so that events in both Eq. (2) and Eq. (4) happen,

3. the call to Lemma 7.7 succeeds.

The success probability is therefore at least

$$\mathbb{P} \left[\text{success} \right] \; \geq \; \frac{\eta^2}{q^{2d}} \cdot \frac{\varepsilon_2}{4\delta} \cdot \frac{1}{5} \; \geq \; \Omega \left(\frac{\varepsilon_2^5}{\delta^3 \cdot q^{2d}} \right) \; = \; \Omega_{q,d}(\varepsilon^5).$$

We have shown that for any $h \in \mathcal{L}((g, \mathcal{J}(\delta) - \varepsilon)$, the algorithm above outputs h with probability at least $\Omega_{q,d}(\varepsilon^5)$. Note that this implicitly proves an upper bound on the list of $\mathcal{O}_{q,d}(1/\varepsilon^5)$.

Therefore, the random choices that the algorithm makes lead it to different elements of the list. We next argue that we can derandomize all random choices in the algorithm, so that all elements of the list can be found with a deterministic algorithm.

1. For the random choice of k, we can try out all possible q^{3d}/η^2 values for k.

26

2. For random (V, β), we can again try out all possible values, which are at most $n^k \cdot 2^k \leq n^{\mathcal{O}_{q,d}(1/\varepsilon^4)}$ in number.

3. Lemma 7.7 can be derandomized using a standard threshold rounding argument, as argued in Lemma 7.9.

Thus, the final algorithm starts with an empty list and goes over all the deterministic choices above. Every $h \in \mathcal{L}(g, \mathcal{J}(\delta) - \varepsilon)$ will be discovered in at least one of these deterministic steps, and we can efficiently check whether $\Delta(g, h) < \mathcal{J}(\delta) - \varepsilon$. If yes, h is added to the output list. ∎

7.2 AEL Code

Let $\mathcal{C}^{AEL}$ be an AEL code determined by an (n, d, λ)-expander graph $G(L, R, E)$, an inner code $\mathcal{C}_0$ of distance δ_0, rate r_0, alphabet size q_0, and an outer code $\mathcal{C}_1$ of distance δ_1, rate r_1 and alphabet size $q_1 = |\mathcal{C}_0|$. The code $\mathcal{C}^{AEL}$ is of alphabet size q_0^d, rate $r_0 r_1$ and (designed) distance $\delta_0 - \frac{\lambda}{\delta_1}$.

Theorem 7.2 (List Decoding AEL codes). *Suppose the code $\mathcal{C}_1$ can be efficiently unique-decoded from radius δ_{dec}. Assume $\lambda \leq \kappa \cdot \delta_{dec} \leq \kappa \cdot \delta_1$, so that the distance of $\mathcal{C}^{AEL}$ is at least $\delta_0 - \kappa$. Then for any $\varepsilon > 0$, the code $\mathcal{C}^{AEL}$ can be list decoded from a radius of $\mathcal{J}(\delta_0 - \kappa) - \varepsilon$ by using $\mathcal{O}_{q,d,\delta_{dec}}(1/\varepsilon^4)$ levels of SoS-hierarchy, in time $n^{\mathcal{O}_{q,d,\delta_{dec}}(1/\varepsilon^4)}$.*

Proof. Let $g \in [q_0^d]^R$ be a received word. Recall that the distance of an AEL codeword h with g is given by $\Delta^R(g, h) = \mathbb{E}_{r \in R}[\mathbf{1}\{g(r) \neq h_{N_R(r)}\}]$.

We again start by applying the algorithmic covering Lemma 4.12 to get a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ of SoS-degree $t \geq 2d(\frac{q^{3d}}{\eta^2} + 2)$ such that for every $h \in \mathcal{L}(g, \mathcal{J}(\delta_0 - \kappa) - \varepsilon)$,

$$\mathbb{E}_r\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}\right]\right] = \Delta^R(\widetilde{\mathbb{E}}[\cdot], h) \leq (\delta_0 - \kappa) - \varepsilon_2$$

Here $\eta > 0$ is a small constant to be chosen later, and $\varepsilon_2 = 2\varepsilon\sqrt{1 - \frac{q^d}{q^d - 1}(\delta_0 - \kappa)} \geq \Omega(\varepsilon)$. Henceforth, we fix an $h \in \mathcal{L}((g, \mathcal{J}(\delta_0 - \kappa) - \varepsilon)$, so that $\Delta^R(\widetilde{\mathbb{E}}[\cdot], h) \leq \delta_0 - \kappa - \varepsilon_2$. Our goal is to recover h.

Pick a random $k \in \{1, \cdots, \lceil \frac{q^{3d}}{\eta^2} \rceil\}$. From Lemma 6.3, we know that with probability at least $\frac{\eta^2}{q^{3d}}$,

$$\mathbb{E}_{v_1, v_2, \cdots, v_k}\left[\mathbb{E}_{\ell, r}\left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(v_1)}, \mathbf{Z}_{N_R(v_2)}, \cdots, \mathbf{Z}_{N_R(v_k)}]\right]\right] \leq \eta$$

We assume that we found a k such that Eq. (1) holds. Let V be the (random) set of k vertices we condition on, that is, $V = \{v_1, v_2, \cdots, v_k\}$, and let $N_R(V) = \cup_{v \in V} N_R(v)$. Then,

$$\mathbb{E}_{V \subseteq R, |V| = k}\left[\mathbb{E}_{\ell, r}\left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(V)}]\right]\right] \leq \eta$$

More explicitly, conditioning on $N_R(V)$ involves sampling an assignment for $N_R(V)$ according to the local distribution of $\mathbf{Z}_{N_R(V)}$. Let this random assignment be β, and we get

$$\mathbb{E}_{\substack{V \subseteq R, |V| = k \\ \beta \sim \mathbf{Z}_{N_R(V)}}}\left[\mathbb{E}_{\ell, r}\left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(V)} = \beta]\right]\right] \leq \eta.$$

By Markov's inequality, and by choosing $\eta = \frac{\varepsilon_2^2 \delta_{dec}}{16(\delta_0 - \kappa)}$,

$$\mathbb{P}_{\substack{V \subseteq R, |V| = k \\ \beta \sim \mathbf{Z}_{N_R(V)}}}\left[\mathbb{E}_{\ell, r}\left[\widetilde{\text{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(V)} = \beta]\right] > \frac{\delta_{dec} \cdot \varepsilon_2}{4}\right] \leq \frac{4\eta}{\delta_{dec} \cdot \varepsilon_2} \leq \frac{\varepsilon_2}{4(\delta_0 - \kappa)} \qquad (7)$$

27

$$\underset{\substack{V \subseteq R, |V|=k \\ \beta \sim \mathbf{Z}_{N_R(V)}}}{\mathbb{P}} \left[\underset{\ell,r}{\mathbb{E}} \left[\widetilde{\mathrm{Cov}}[\mathbf{Z}_{N_L(\ell)}, \mathbf{Z}_{N_R(r)} | \mathbf{Z}_{N_R(V)} = \beta] \right] \leq \frac{\delta_{dec} \cdot \varepsilon_2}{4} \right] \geq 1 - \frac{\varepsilon_2}{4(\delta_0 - \kappa)} \tag{8}$$

As in the Tanner case, we next claim that the distance is preserved with significant probability when conditioning randomly. Let $h \in \mathcal{L}(g, \mathcal{J}(\delta_0 - \kappa) - \varepsilon)$ so that $\Delta^R(\widetilde{\mathbb{E}}[\cdot], h) < \delta_0 - \kappa - \varepsilon_2$. Using a similar argument as in the Tanner case,

$$\underset{(V,\beta)}{\mathbb{E}} \left[\Delta^R \left(\widetilde{\mathbb{E}} \left[\cdot \, | \mathbf{Z}_{N_R(V)} = \beta \right], h \right) \right] \leq (\delta_0 - \kappa) - \varepsilon_2 \tag{9}$$

which allows us to claim via Markov's inequality that

$$\underset{(V,\beta)}{\mathbb{P}} \left[\Delta^R \left(\widetilde{\mathbb{E}} \left[\cdot \, | \mathbf{Z}_{N_R(V)} = \beta \right], h \right) \leq (\delta_0 - \kappa) - \frac{\varepsilon_2}{2} \right] \geq \frac{\varepsilon_2/2}{(\delta_0 - \kappa) - \varepsilon_2 + \varepsilon_2/2} \geq \frac{\varepsilon_2}{2(\delta_0 - \kappa)} \tag{10}$$

Again, let (V, β) be a conditioning such that events in both Eq. (8) and Eq. (10) hold (which happens with probability at least $\frac{\varepsilon_2}{4(\delta_0 - \kappa)}$). Let $\widetilde{\mathbb{E}}^*[\cdot] = \widetilde{\mathbb{E}}[\cdot | \mathbf{Z}_{N_R(V)} = \beta]$ be an η-good pseudocodeword that satisfies Eq. (10). This means

$$\delta_0 - \frac{\lambda + \eta}{\Delta^L(\widetilde{\mathbb{E}}^*[\cdot], h)} \leq \Delta^R(\widetilde{\mathbb{E}}^*[\cdot], h) \leq (\delta_0 - \kappa) - \varepsilon_2/2$$

Rearranging, we get that

$$\kappa + \varepsilon_2/2 \leq \frac{\lambda + \eta}{\Delta^L(\widetilde{\mathbb{E}}^*[\cdot], h)} \leq \frac{\lambda + \delta_{dec} \cdot \varepsilon_2/4}{\Delta^L(\widetilde{\mathbb{E}}^*[\cdot], h)} \leq \frac{\kappa \cdot \delta_{dec} + \delta_{dec} \cdot \varepsilon_2/4}{\Delta^L(\widetilde{\mathbb{E}}^*[\cdot], h)},$$

which gives the required bound on $\Delta^L(\widetilde{\mathbb{E}}^*[\cdot], h)$ as

$$\Delta^L(\widetilde{\mathbb{E}}^*[\cdot], h) \leq \delta_{dec} - \frac{\delta_{dec}\varepsilon_2/4}{\kappa + \varepsilon_2/2} \leq \delta_{dec} - \frac{\delta_{dec}\varepsilon_2}{4\delta_0}$$

Finally, we use Lemma 7.8 to find h using $\widetilde{\mathbb{E}}^*[\cdot]$ with probability at least $\varepsilon_2/4\delta_0$. The final success probability is at least

$$\frac{\eta^2}{q^{3d}} \cdot \frac{\varepsilon_2}{4(\delta_0 - \kappa)} \cdot \frac{\varepsilon_2}{4\delta_0} \geq \Omega\left(\frac{\varepsilon_2^6 \delta_{dec}^2}{q^{3d} \delta_0^4} \right) \geq \Omega_{q,d,\delta_{dec}}(\varepsilon^6).$$

Just as in the case of Tanner codes, this algorithm can be derandomized by trying out all possible random choices made by the algorithm, to give a deterministic algorithm that recovers the list. ∎

Note that while the theorem above deals with list decoding, it can be easily adapted for list recovery by replacing the use of Lemma 4.12 by Lemma 4.13. Next, we use the AEL amplification scheme to construct near-MDS codes list decodable up to the Johnson bound.

Theorem 7.3. *For any $1/2 > \varepsilon_1, \varepsilon_2 > 0$, there is an infinite family of codes $\mathcal{C}$ of blocklength n with the following properties:*

(i) The rate of the code is ρ and distance is at least $1 - \rho - \varepsilon_1$.

(ii) The code is over an alphabet of size $2^{\mathcal{O}(\varepsilon_1^{-6} \log(1/\varepsilon_1))}$.

(iii) The code can be list decoded from radius $\mathcal{J}(1 - \rho - \varepsilon_1) - \varepsilon_2$ in time $n^{\mathcal{O}_{\varepsilon_1}(1/\varepsilon_2^4)}$.

Proof. We sketch how to instantiate Theorem 7.2 to obtain such codes.

Suppose we are working with (n, d, λ)-expander.

Choose the inner code $\mathcal{C}_0$ to be a Reed Solomon code of rate ρ_0, distance $1 - \rho_0$ and alphabet size $q_0 = d$ (or any MDS code). Choose the outer code $\mathcal{C}_1$ over alphabet of size $|\mathcal{C}_0| = d^{\rho_0 \cdot d}$ to have rate $1 - \varepsilon_1$ that can be unique decoded from radius $\delta_{dec} = \Omega(\varepsilon_1^2)$, such as the one constructed in [GI05].

Let $\lambda = \kappa \delta_{dec}$ with $\kappa = \varepsilon_1$, so that $\lambda \leq \Theta(\varepsilon_1^3)$ and $d = \Theta(1/\varepsilon_1^6)$.

The rate of the final AEL code is $\rho := (1 - \varepsilon_1)\rho_0$, and the distance is at least

$$
\begin{aligned}
(1 - \rho_0) - \frac{\lambda}{\delta_1} &\geq 1 - \frac{\rho}{1 - \varepsilon_1} - \frac{\lambda}{\delta_{dec}} \\
&\geq 1 - \rho - 2\varepsilon_1 \rho - \varepsilon_1 \\
&\geq 1 - \rho - 3\varepsilon_1
\end{aligned}
$$

The alphabet size is $q_0^d = d^d = 2^{\mathcal{O}\left(\varepsilon_1^{-6} \log(1/\varepsilon_1)\right)}$.

For list decodability, we use Theorem 7.2 to claim that the above code can be list decoded from $\mathcal{J}(1 - \rho - 3\varepsilon_1) - \varepsilon_2$ in time $n^{\mathcal{O}_{q_0, d, \delta_{dec}}(1/\varepsilon_2^4)} = n^{\mathcal{O}_{\varepsilon_1}(1/\varepsilon_2^4)}$.

Replace ε_1 by $\varepsilon_1/3$ to get the final result. ∎

Note that we can also deal with $\mathcal{C}_1$ that can be decoded from smaller radius like $\mathcal{O}(\varepsilon_1^3)$, by suitably adjusting λ and paying the cost in alphabet size. Above, we have not chosen parameters optimally to keep the exposition simple. The alphabet size in the code constructed in [GI05] was smaller than what we ask for here, but alphabet size can always be increased while preserving rate, distance and unique decoding radius by folding multiple symbols together into one. This looks like multiple symbols of the outer code being assigned to the same left node in the AEL construction.

7.3 Decoding from fractional vectors

This section has auxiliary claims needed to finish the list decoding algorithm proof. In the Tanner code case, we reduce to unique decoding of the same code from an arbitrarily small radius. In AEL, we reduce to unique decoding of the base (outer) code. In both these cases, we have some ε slack, that is sufficient for randomized rounding to produce a (corrupted) word within the unique decoding radius.

Lemma 7.4. *Let $\mathcal{C}$ be an $[n, \delta, \rho]_q$ code, which is unique decodable from distance $\delta_{dec} \leq \delta/2$ in time $\mathcal{T}(n)$. Given a collection of distributions $\mathcal{D} = \{\mathcal{D}_i\}_{i \in [n]}$, each of them supported on $[q]$, there is a unique codeword h that satisfies*

$$
\mathbb{E}_i \left[\mathbb{E}_{j \sim \mathcal{D}_i} [\mathbf{1}\{h_i \neq j\}] \right] \leq \delta_{dec} - \varepsilon
$$

This codeword h can be found in time $\mathcal{O}(qn) + \mathcal{T}(n)$ with probability at least ε/δ_{dec}.

Proof. First, we show uniqueness of h. Let h, h' be two codewords in $\mathcal{C}$ such that

$$
\mathbb{E}_i \left[\mathbb{E}_{j \sim \mathcal{D}_i} [\mathbf{1}\{h_i \neq j\}] \right] \leq \delta_{dec} - \varepsilon \qquad \text{and} \qquad \mathbb{E}_i \left[\mathbb{E}_{j \sim \mathcal{D}_i} [\mathbf{1}\{h'_i \neq j\}] \right] \leq \delta_{dec} - \varepsilon
$$

For any $j \in [q]$, we have $\mathbf{1}\{h_i \neq h'_i\} \leq \mathbf{1}\{h_i \neq j\} + \mathbf{1}\{h'_i \neq j\}$. Therefore,

$$
\begin{aligned}
\Delta(h, h') = \mathbb{E}_i [\mathbf{1}\{h_i \neq h'_i\}] &\leq \mathbb{E}_i \left[\mathbb{E}_{j \sim \mathcal{D}_i} [\mathbf{1}\{h_i \neq j\} + \mathbf{1}\{h'_i \neq j\}] \right] \\
&= \mathbb{E}_i \left[\mathbb{E}_{j \sim \mathcal{D}_i} [\mathbf{1}\{h_i \neq j\}] \right] + \mathbb{E}_i \left[\mathbb{E}_{j \in \mathcal{D}_i} [\mathbf{1}\{h'_i \neq j\}] \right] \leq 2\delta_{dec} - 2\varepsilon < \delta.
\end{aligned}
$$

29

By distance property of the code, this means that $\Delta(h, h') = 0$, or that $h = h'$.

The algorithm to find h is to independently sample from every distribution to get a random $g \in [q]^n$, and then issue a unique decoding call from g. We show that with significant probability, h lies in the δ_{dec} radius ball around g, which will show that algorithm succeeds with that probability.

$$\mathbb{E}_g \left[\Delta(g, h) \right] = \mathbb{E}_g \left[\mathbb{E}_i \left[\mathbf{1}\{g_i \neq h_i\} \right] \right] = \mathbb{E}_i \left[\mathbb{E}_g \left[\mathbf{1}\{g_i \neq h_i\} \right] \right] = \mathbb{E}_i \left[\mathbb{E}_{g_i \sim \mathcal{D}_i} \left[\mathbf{1}\{g_i \neq h_i\} \right] \right] \leq \delta_{dec} - \varepsilon.$$

Thus, by Markov's inequality, we have, $\mathbb{P}_g[\Delta(g, h) \leq \delta_{dec}] \geq \frac{\varepsilon}{\delta_{dec}}$, which proves the claim. ■

Remark 7.5. *The success probability in the Lemma 7.4 can be amplified by repeated sampling. Moreover, the fact that we reduce to the unique decoding algorithm of $\mathcal{C}$ is not important, as it is also possible to use a list decoding algorithm for $\mathcal{C}$ from distance δ_{dec} to find h, as long as the sampled g satisfies $\delta(g, h) < \delta_{dec}$. In that case, we output a random element of list obtained, which incurs an additional loss of $1/L$ factor, where L is the list size guaranteed by list decoding algorithm for $\mathcal{C}$ up to δ_{dec}.*

Remark 7.6. *As shown in Lemma 7.9, this argument can be derandomized using threshold rounding. The use of Lemma 7.4 in the next two lemmas can therefore be replaced by Lemma 7.9.*

Next, we use pseudocodewords that lie in the unique decoding ball to construct the collection of distributions needed by Lemma 7.4, for the Tanner and AEL cases.

Lemma 7.7 (Unique decoding from Tanner pseudocodewords). *Let $\mathcal{C}^{Tan}$ be a code as in Section 7.1, with distance $\delta = \delta_0(\delta_0 - \lambda)$ and $\lambda \leq \delta_0/3$, and in particular it can be unique decoded from radius $\delta/4$ in time $\mathcal{O}(|E|)$. Given a Tanner pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ such that $\Delta(\widetilde{\mathbb{E}}[\cdot], h) < \delta/5$, we can find h in time $\mathcal{O}(|E|)$ with probability at least $1/5$.*

Proof. The pseudocodeword gives a collection of distributions $\mathcal{D} = \{\mathcal{D}_e\}_{e \in E}$, each distribution over $[q]$. The eth distribution $\mathcal{D}_e$ gives a weight of $\widetilde{\mathbb{E}}[\mathbf{1}\{Z_e = j\}]$ to the value $j \in [q]$.

The distance $\Delta(\widetilde{\mathbb{E}}[\cdot], h)$ translates to

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) = \mathbb{E}_e \left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{Z_e \neq h_e\} \right] \right]$$

$$= \mathbb{E}_e \left[\widetilde{\mathbb{E}} \left[\sum_{j \in [q]} \mathbf{1}\{Z_e = j\} \mathbf{1}\{h_e \neq j\} \right] \right]$$

$$= \mathbb{E}_e \left[\sum_{j \in [q]} \mathbf{1}\{h_e \neq j\} \widetilde{\mathbb{E}} \left[\mathbf{1}\{Z_e = j\} \right] \right]$$

$$= \mathbb{E}_e \left[\mathbb{E}_{j \sim \mathcal{D}_e} \left[\mathbf{1}\{h_e \neq j\} \right] \right]$$

We can therefore use this collection of distributions to obtain a codeword $h \in \mathcal{C}^{Tan}$ via Lemma 7.4 with $\delta_{dec} = \delta/4$ and $\varepsilon = \delta/20$ such that $\Delta(\widetilde{\mathbb{E}}[\cdot], h) < \delta/5$. ■

Next, we use similar ideas to round and decode from AEL pseudocodewords. We borrow the terminology used for AEL codes from Section 7.2.

Lemma 7.8 (Unique decoding from AEL pseudocodewords). *Let $\mathcal{C}^{AEL}$ be a code as in Section 7.2, with $\lambda \leq \kappa \cdot \delta_{dec}$ and distance at least $\delta_0 - \kappa$. Assume that the outer code $\mathcal{C}_1$ can be unique decoded from radius δ_{dec} in time $\mathcal{T}(n)$. Given an AEL pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ and $h \in \mathcal{C}^{AEL}$ such that $\Delta^L(\widetilde{\mathbb{E}}[\cdot], h) \leq \delta_{dec} - \varepsilon$, we can find h in time $\mathcal{O}(n) + \mathcal{T}(n)$ with probability at least ε/δ_{dec}.*

Proof. First, we use the given pseudocodewords to build a collection of distributions $\mathcal{D} = \{\mathcal{D}_\ell\}_{\ell \in L}$, each distribution over $[q_1]$. Recall that $\mathcal{C}_0$ can be seen as a map from $[q_1]$ to $[q_0]^d$. The ℓth distribution gives a weight of $\widetilde{\mathbb{E}}[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} = \mathcal{C}_0(\alpha)\}]$ to $\alpha \in [q_1]$.

Let $\overline{h}$ be the codeword in $\mathcal{C}_1$ corresponding to the codeword h. That is, $\overline{h}$ is such that $\mathcal{C}_0(\overline{h}_\ell) = h_{N_L(\ell)}$. With the collection of distributions $\mathcal{D}$ defined, we relate $\Delta^L(\widetilde{\mathbb{E}}[\cdot], h)$ to agreement of $\overline{h}$ with $\mathcal{D}$.

$$\Delta^L(\widetilde{\mathbb{E}}[\cdot], h) = \mathbb{E}_\ell \left[\widetilde{\mathbb{E}} \left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\} \right] \right]$$

$$= \mathbb{E}_\ell \left[\widetilde{\mathbb{E}} \left[\sum_{\alpha \in [q_1]} \mathbf{1}\{\mathbf{Z}_{N_L(\ell)} = \mathcal{C}_0(\alpha)\} \mathbf{1}\{h_{N_L(\ell)} \neq \mathcal{C}(\alpha)\} \right] \right]$$

$$= \mathbb{E}_\ell \left[\sum_{\alpha \in [q_1]} \mathbf{1}\{h_{N_L(\ell)} \neq \mathcal{C}_0(\alpha)\} \widetilde{\mathbb{E}} \left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} = \mathcal{C}_0(\alpha)\} \right] \right]$$

$$= \mathbb{E}_\ell \left[\mathbb{E}_{\alpha \sim \mathcal{D}_\ell} \left[\mathbf{1}\{h_{N_L(\ell)} \neq \mathcal{C}_0(\alpha)\} \right] \right]$$

$$= \mathbb{E}_\ell \left[\mathbb{E}_{\alpha \sim \mathcal{D}_\ell} \left[\mathbf{1}\{\overline{h}_\ell \neq \alpha\} \right] \right]$$

We call Lemma 7.4 for the code $\mathcal{C}_1$ with the collection of distributions $\mathcal{D}$ and find $\overline{h}$, and therefore h, with probability at least ε/δ_{dec}. ∎

To end this section, we note that the rounding from fractional vectors above can be derandomized through a standard method known as threshold rounding. Similar ideas are used to derandomize the classical Generalized Minimum Distance decoding for concatenated codes.

Lemma 7.9. *Let $\mathcal{C}$ be an $[n, \delta, \rho]_q$ code, which is unique decodable from distance $\delta_{dec} \leq \delta/2$ in time $\mathcal{T}(n)$. Given a collection of distributions $\mathcal{D} = \{\mathcal{D}_i\}_{i \in [n]}$, each of them supported on $[q]$ described as a collection q weights that sum to 1, there is a unique codeword h that satisfies*

$$\mathbb{E}_i \left[\mathbb{E}_{j \sim \mathcal{D}_i} [\mathbf{1}\{h_i \neq j\}] \right] \leq \delta_{dec}$$

This codeword h can be found in time $\mathcal{O}(qn) \cdot \mathcal{T}(n)$ with a deterministic algorithm.

Proof. The uniqueness of h is as in proof of Lemma 7.4.

Let the weight on $j \in [q]$ according to $\mathcal{D}_i$ be w_{ij}, so that $\sum_{j \in [q]} w_{ij} = 1$. We replace the randomized rounding of Lemma 7.4 by the following process:

(i) Define the cumulative sums $c_{ij} = \sum_{j' \leq j} w_{ij'}$, so that $c_{iq} = 1$. Define $c_{i0} = 0$.

(ii) For each $i \in [n]$, embed $\mathcal{D}_i$ into the interval $[0, 1]$ as $q + 1$ points $(c_{i0} = 0, c_{i1}, c_{i2}, \cdots, c_{iq} = 1)$.

(iii) Choose $\theta \in [0, 1]$ uniformly at random.

(iv) Build $h' \in [q]^n$ coordinate-wise as follows. For $i \in [n]$, if

$$\theta \in \left[c_{i(j-1)}, c_{ij} \right),$$

then $h'_i = j$. This ensures that $\mathbb{P}_\theta[h'_i = j] = w_{ij}$.

We show that h' has the same distance from h in expectation as the collection of distributions $\{\mathcal{D}_i\}_{i\in[n]}$.

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\theta\in[0,1]}\left[\Delta(h',h)\right] &= \mathop{\mathbb{E}}_{\theta\in[0,1]}\left[\mathop{\mathbb{E}}_{i}\left[\mathbf{1}\{h_i' \neq h_i\}\right]\right] \\
&= \mathop{\mathbb{E}}_{\theta\in[0,1]}\left[\mathop{\mathbb{E}}_{i}\left[\mathbf{1}\{\theta \notin [c_{i(h_i-1)}, c_{ih_i})\}\right]\right] \\
&= \mathop{\mathbb{E}}_{i}\left[\mathop{\mathbb{E}}_{\theta\in[0,1]}\left[\mathbf{1}\{\theta \notin [c_{i(h_i-1)}, c_{ih_i})\}\right]\right] \\
&= \mathop{\mathbb{E}}_{i}\left[1 - w_{ih_i}\right] \\
&= \mathop{\mathbb{E}}_{i}\left[\mathop{\mathbb{E}}_{j\sim\mathcal{D}_i}\left[\mathbf{1}\{h_i \neq j\}\right]\right] \leq \delta_{dec}
\end{aligned}
$$

Therefore, rounding according to a random threshold θ produced an h' that is at most δ_{dec} distance away from h. As the final step to derandomization, note that two thresholds θ_1, θ_2 produce the exact same h' if there is no point from step (ii) above embedded between θ_1 and θ_2. Total number of points embedded is at most $q \cdot n$, and so it suffices to try at most $\mathcal{O}(q \cdot n)$ many thresholds to produce all the different h' possible - one of which must be δ_{dec}-close to h. ■

8 List Decoding Concatenated Codes

In this section, we will adapt the techniques developed earlier to decode concatenated codes. Concatenation is a useful operation to obtain codes with smaller alphabet size starting from a code over large alphabet. Previous works on list decoding of concatenated codes [GS00, GR06, GS02] seem to all rely on list recovery of outer code, with intricate weights to be passed along with inner codewords. We will only use list decodability of outer code.

First, we show that the covering lemma based argument can be used to decode the concatenated code up to the Johnson radius corresponding to product of decoding radius of outer code and the distance of inner code. That is, if the distance of inner code is δ_0, distance of outer code is δ_1, and the decoding radius of outer code is δ_{dec}, we can decode the concatenated code up to radius $\mathcal{J}(\delta_{dec} \cdot \delta_0)$.

Moreover, since concatenated codes do not involve expansion for their distance proof, we do not need to deal with SoS-based pseudocodewords or any low-covariance conditions. In fact, our pseudocodewords will just be local distributions over the inner code for each coordinate of the outer code. Since this set of pseudocodewords can be described as the feasible set corresponding to linear constraints over $\mathcal{O}(n)$ variables, we can minimize the appropriate norm (which is a convex function) via Ellipsoid method to get a pseudocodeword with the covering property in time $n^{\mathcal{O}(1)}$.

Note that this is weaker than decoding up to $\mathcal{J}(\delta_1 \cdot \delta_0)$, which is the Johnson radius corresponding to the true distance of the concatenated code. In Section 8.2, we will see that by using the decoder of outer code in a white box way, we can get to the Johnson bound $\mathcal{J}(\delta_1 \cdot \delta_0)$ when the outer code supports list decoding through our Covering Lemma based machinery, like in the case of near-MDS codes of Theorem 7.3.

8.1 List decoding arbitrary concatenated codes

Let the $[n, \delta_1, \rho_1]_{q_1}$ outer code be $\mathcal{C}_1$ and the $[d, \delta_0, \rho_0]_{q_0}$ inner code be $\mathcal{C}_0$, with $q_1 = q_0^{\rho_0 \cdot d}$. Let the concatenated code be $\mathcal{C}^*$ with distance at least $\delta = \delta_0 \cdot \delta_1$. A codeword $h^* = \mathcal{C}_0(h)$ of the concatenated code can be seen as a tuple $h^* = (h_1^*, h_2^*, \cdots, h_n^*)$ where each $h_i^* \in \mathcal{C}_0$, or $h_i^* = \mathcal{C}_0(h_i)$ for some $h_i \in [q_1]$. Note that not all tuples of this form are codewords, and $(\mathcal{C}_0(f_1), \mathcal{C}_0(f_2), \cdots, \mathcal{C}_0(f_n)) \in \mathcal{C}^*$ iff $(f_1, f_2, \cdots, f_n) \in \mathcal{C}_1$.

Definition 8.1. *A pseudocodeword of the concatenated code is a psuedoexpectation operator $\widetilde{\mathbb{E}}[\cdot]$ of degree d over the variables $\mathbf{Z} = \left\{ Z_{i,j,k} \right\}_{i \in [n], j \in [d], k \in [q_0]}$ that respects the following d-local constraints:*

    (i) $Z_{i,j,k}^2 = Z_{i,j,k}$

    (ii) For every $i \in [n], j \in [d]$, $\sum_{k \in [q_0]} Z_{i,j,k} = 1$

    (iii) $\forall i \in [n], \quad (\mathbf{Z}_{i,1}, \mathbf{Z}_{i,2}, \cdots, \mathbf{Z}_{i,d}) \in \mathcal{C}_0$.

We no longer enforce the constraint for non-negativity of squares of polynomials, and in fact, our pseudoexpectation operators are just a collection of n distributions $\{\mathcal{D}_i\}_{i \in [n]}$ over $[q_1]$. The weight assigned to $f \in [q_1]$ in distribution $\mathcal{D}_i$ is $\widetilde{\mathbb{E}}[\mathbf{1}\{\mathbf{Z}_i = \mathcal{C}_0(f)\}]$.

Following is the natural generalization of distances to pseudocodewords.

Definition 8.2 (Distance from a pseudocodeword). *The distance of a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ and a codeword h^* is defined as*

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h^*) = \mathop{\mathbb{E}}_{i \in [n]} \left[\widetilde{\mathbb{E}}\left[\Delta(\mathbf{Z}_i, h_i^*) \right] \right]$$

Lemma 8.3. *Assume that $\mathcal{C}_1$ can be list-decoded from radius δ_{dec} in time $\mathcal{T}(n)$ with list size L.*

For any $\varepsilon > 0$, there is a deterministic algorithm that given a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ outputs the list

$$\mathcal{L}(\widetilde{\mathbb{E}}[\cdot], \delta_0 \cdot \delta_{dec}) := \left\{ h^* \in \mathcal{C}^* \; : \; \Delta(h^*, \widetilde{\mathbb{E}}[\cdot]) < \delta_0 \cdot \delta_{dec} \right\}$$

and runs in time $2^{\mathcal{O}(d)} n^{\mathcal{O}(1)} + \mathcal{O}(2^d \cdot n) \cdot \mathcal{T}(n)$.

Proof. We use $\widetilde{\mathbb{E}}[\cdot]$ to get local distributions $\mathcal{D}_i$ over $[q_1]$ for each coordinate i.

$$\begin{aligned}
\Delta(\widetilde{\mathbb{E}}[\cdot], h^*) = \mathop{\mathbb{E}}_{i \in [n]} \left[\widetilde{\mathbb{E}}\left[\Delta(\mathbf{Z}_i, h_i^*) \right] \right] &= \mathop{\mathbb{E}}_{i \in [n]} \left[\mathop{\mathbb{E}}_{f \sim \mathcal{D}_i} \left[\Delta\left(\mathcal{C}_0(f), h_i^* \right) \right] \right] \\
&\geq \mathop{\mathbb{E}}_{i \in [n]} \left[\mathop{\mathbb{E}}_{f \sim \mathcal{D}_i} \left[\mathbf{1}\{f \neq h_i\} \Delta\left(\mathcal{C}_0(f), h_i^* \right) \right] \right] \\
&\geq \mathop{\mathbb{E}}_{i \in [n]} \left[\mathop{\mathbb{E}}_{f \sim \mathcal{D}_i} \left[\mathbf{1}\{f \neq h_i\} \delta_0 \right] \right] \\
&\geq \mathop{\mathbb{E}}_{i \in [n]} \left[\mathop{\mathbb{E}}_{f \sim \mathcal{D}_i} \left[\mathbf{1}\{f \neq h_i\} \Delta\left(\mathcal{C}_0(f), h_i^* \right) \right] \right] \\
&= \delta_0 \cdot \mathop{\mathbb{E}}_{i \in [n]} \left[\mathop{\mathbb{E}}_{f \sim \mathcal{D}_i} \left[\mathbf{1}\{f \neq h_i\} \right] \right]
\end{aligned}$$

Any codeword $h^* = \mathcal{C}_0(h) \in \mathcal{L}$ must have the property that $\Delta(\widetilde{\mathbb{E}}[\cdot], h^*) < \delta_{dec} \cdot \delta_0$, so that

$$\mathop{\mathbb{E}}_{i} \left[\mathop{\mathbb{E}}_{f \sim \mathcal{D}_i} \left[\mathbf{1}\{f \neq h_i\} \right] \right] < \delta_{dec}.$$

Finally, we use Lemma 7.9 to decode from the above agreement. The only modification is that since we might potentially deal with the *list* decoding algorithm of outer code $\mathcal{C}_1$, we take a union of all the lists generated by the different calls corresponding to different thresholds, and then prune it finally. For any $h^* = \mathcal{C}_0(h) \in \mathcal{L}$, there is some threshold for which the $[q_1]^n$ string generated in Lemma 7.9 will be at distance $< \delta_{dec}$ from h. Therefore, h will be contained in at least one of the lists discovered by the algorithm. $\blacksquare$

Lemma 8.4. *For any $\varepsilon > 0$, there is an algorithm that given $g \in [q_0]^{nd}$ and $\delta > 0$, finds a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ such that $\mathcal{L}(g, \mathcal{J}(\delta) - \varepsilon) \subseteq \mathcal{L}(\widetilde{\mathbb{E}}[\cdot], \delta - \varepsilon_2)$ in time $2^{\mathcal{O}(d)} \cdot n^{\mathcal{O}(1)}$ for some $\varepsilon_2 > 0$.*

Proof Sketch. This is again an algorithmic implementation of the covering lemma, with the distributions over codewords to be relaxed to distributions over pseudocodewords as defined above. Minimizing the appropriate norm while optimizing over the convex set of pseudocodewords gives us this covering property. We omit the details since the argument is very similar to Lemma 4.7.

Since the above optimization problem minimizes a convex quadratic function with linear constraints on at most $2^d \cdot n$ variables, we can get a running time of $2^{\mathcal{O}(d)} \cdot n^{\mathcal{O}(1)}$. ■

Theorem 8.5. *Let $\mathcal{C}_0$ be a binary inner code of blocklength d, distance δ_0 and rate ρ_0. Also let $\mathcal{C}_1$ be an outer code of blocklength n, distance δ_1 and rate ρ_1 on an alphabet of size $|\mathcal{C}_0| = 2^{\rho_0 d}$. Assume that $\mathcal{C}_1$ can be list-decoded from radius δ_{dec} in time $\mathcal{T}(n)$ with list size L.*

Then the code $\mathcal{C}^*$ obtained by concatenating $\mathcal{C}_1$ with $\mathcal{C}_0$ can be list decoded up to a radius of $\mathcal{J}(\delta_{dec}\delta_0)$ in time $2^{\mathcal{O}(d)} n^{\mathcal{O}(1)} + \mathcal{O}(2^d \cdot n) \cdot \mathcal{T}(n)$.

Proof. Given g such that we wish to find $\mathcal{L} = \mathcal{L}(g, \mathcal{J}(\delta_{dec} \cdot \delta_0) - \varepsilon)$, we first use Lemma 8.4 to find a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ that covers the list, and then use it via Lemma 8.3 to find $\mathcal{L}$. ■

8.2 List Decoding with outer AEL codes

In this section, we show that when the near-MDS codes obtained using AEL distance amplification are used for concatenation with smaller alphabet codes, we can list decode the smaller alphabet code up to its Johnson bound. This will not be done by a black-box call to the list decoding/list recovery algorithm of the outer code, but we will crucially use the SoS-based list decoding strategy for the outer code.

To the best of our knowledge, list-decoding to the Johnson bound of the concatenated code has not been achieved for the Reed-Solomon outer code. This shows that while our near-MDS codes via AEL construction match Reed-Solomon codes in terms of list decoding radius, they have some extra desirable features. Of course, the runtime of our algorithms is quite poor compared to the near-linear time Reed-Solomon decoders.

The pseudocodewords for the concatenated code will be concatenations of the pseudocodewords of outer AEL code. Since the covering lemma works irrespective of the code we are working with, we can get a cover for the list of (final) codewords by efficiently optimizing over the pseudocodewords of outer code. Then a simple argument shows that the distance property for outer pseudocodewords translates to distance property for concatenated pseudocodewords.

Let us recall some notation for AEL Codes. Let $\mathcal{C}^{AEL}$ be an AEL code determined by an (n, d, λ)-expander graph $G = (L, R, E)$, an inner code $\mathcal{C}_0$ of distance δ_0, rate r_0, alphabet size q_0, and an outer code $\mathcal{C}_1$ of distance δ_1, rate r_1 and alphabet size $q_1 = |\mathcal{C}_0|$. The code $\mathcal{C}^{AEL}$ is of alphabet size q_0^d, rate $r_0 r_1$ and (designed) distance $\delta_0 - \frac{\lambda}{\delta_1}$.

Suppose we concatenate $\mathcal{C}^{AEL}$ with an $[d_2, \delta_2, r_2]_{q_2}$ code $\mathcal{C}_2$ such that $q_0^d = q_2^{r_2 \cdot d_2}$, to obtain a new $\left[nd_2, \left(\delta_0 - \frac{\lambda}{\delta_1} \right) \delta_2, r_0 r_1 r_2 \right]_{q_2}$ code $\mathcal{C}_3$. Note that for the code $\mathcal{C}_3$, distance is defined for $f_1, f_2 \in [q_2]^{nd_2}$ as

$$\Delta^*(f_1, f_2) = \mathop{\mathbb{E}}_{r \in R, j \in [d_2]} [\mathbf{1}\{f_1(r, j) \neq f_2(r, j)\}] \tag{11}$$

which can also be extended to distances between a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ and a codeword $h^* \in \mathcal{C}_3$ as

$$\Delta^*(\widetilde{\mathbb{E}}[\cdot], h^*) = \mathop{\mathbb{E}}_{r \in R, j \in [d_2]} \left[\widetilde{\mathbb{E}} \left[\mathbf{1}\{\mathcal{C}_2(\mathbf{Z}_{N_R(r)})(j) \neq h^*(r, j)\} \right] \right]$$

If the codeword $h^* \in \mathcal{C}_3$ is obtained by concatenating $h \in \mathcal{C}^{AEL}$ with $\mathcal{C}_2$, we denote $h^* = \mathcal{C}_2(h)$, and the above distance expression is the same as

$$\Delta^*(\widetilde{\mathbb{E}}[\cdot], \mathcal{C}_2(h)) = \mathop{\mathbb{E}}_{r \in R, j \in [d_2]} \left[\widetilde{\mathbb{E}} \left[\mathbf{1}\{\mathcal{C}_2(\mathbf{Z}_{N_R(r)})(j) \neq \mathcal{C}_2(h_{N_R(r)})(j)\} \right] \right]$$

Theorem 8.6. *Let the code $\mathcal{C}_1$ be decodable from radius δ_{dec}, and $\lambda \leq \kappa \cdot \delta_{dec}$, so that the distance of code $\mathcal{C}_3$ is at least $\delta_2(\delta_0 - \kappa)$. For every $\varepsilon > 0$, the code $\mathcal{C}_3$ can be list decoded up to the radius $\mathcal{J}_{q_2}\left((\delta_0 - \kappa)\,\delta_2\right) - \varepsilon$ in time $n^{\mathcal{O}_{q,d,\delta_{dec}}(1/\varepsilon^4)}$.*

Proof Sketch. We consider the same pseudocodewords as we did for AEL codes, and recall that we proved the following distance property for η-good pseudocodewords in Lemma 5.7. For any $h \in \mathcal{C}^{AEL}$,

$$\mathop{\mathbb{E}}_{r}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}\right]\right] \geq \delta_0 - \frac{\lambda + \eta}{\mathbb{E}_\ell[\widetilde{\mathbb{E}}[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}]]}$$

We extend the above distance property to the distance according to final $\mathcal{C}_3$,

$$\mathop{\mathbb{E}}_{r \in R, j \in [d_2]}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathcal{C}_2(\mathbf{Z}_{N_R(r)})(j) \neq \mathcal{C}_2(h_{N_R(r)})(j)\}\right]\right] = \mathop{\mathbb{E}}_{r}\left[\widetilde{\mathbb{E}}\left[\Delta\left(\mathcal{C}_2(\mathbf{Z}_{N_R(r)}), \mathcal{C}_2(h_{N_R(r)})\right)\right]\right]$$

$$\geq \delta_2 \cdot \mathop{\mathbb{E}}_{r}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_R(r)} \neq h_{N_R(r)}\}\right]\right]$$

$$\geq \delta_2\left(\delta_0 - \frac{\lambda + \eta}{\mathbb{E}_\ell[\widetilde{\mathbb{E}}[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}]]}\right)$$

That is,

$$\Delta^*\left(\widetilde{\mathbb{E}}\left[\cdot\right], \mathcal{C}_2(h)\right) \geq \delta_2\left(\delta_0 - \frac{\lambda + \eta}{\Delta^L\left(\widetilde{\mathbb{E}}[\cdot], h\right)}\right) \tag{12}$$

The key distance property established, the rest of the argument is same as while decoding AEL codes. Given g to be decoded, let's call the list of codewords at distance less than $\mathcal{J}(\delta_2(\delta_0 - \kappa)) - \varepsilon$ as $\mathcal{L}$. We find a pseudocodeword $\widetilde{\mathbb{E}}[\cdot]$ that is close to all the codewords in $\mathcal{L}$. For any $h^* \in \mathcal{L}$ such that $h^* = \mathcal{C}_2(h)$,

$$\Delta^*(\widetilde{\mathbb{E}}\left[\cdot\right], h^*) = \Delta^*(\widetilde{\mathbb{E}}\left[\cdot\right], \mathcal{C}_2(h)) < \delta_2(\delta_0 - \kappa) - \varepsilon_2$$

where $\varepsilon_2 = 2\sqrt{1 - \frac{q_2}{q_2-1}\delta_2(\delta_0 - \kappa)} \cdot \varepsilon = \Theta(\varepsilon)$.

The covering lemma used here minimizes ℓ_2-norm of the embedding corresponding to the concatenated code $\mathcal{C}_3$ with the alphabet size q_2, while the relaxation for pseudocodeword was defined for $\mathcal{C}^{AEL}$ which does not depend on $\mathcal{C}_2$ at all. However, this embedding according to $\mathcal{C}_3$ is a linear function of $\widetilde{\mathbb{E}}[\cdot]$, therefore it is still a convex function that is minimized.

By conditioning $\widetilde{\mathbb{E}}[\cdot]$, we obtain another pseudocodeword $\widetilde{\mathbb{E}}^*[\cdot]$ which is η-good and retains its closeness to $\mathcal{C}_2(h)$. By choosing λ and η small enough, Eq. (12) allows us to conclude that

$$\Delta^L\left(\widetilde{\mathbb{E}}\left[\cdot\right], h\right) = \mathop{\mathbb{E}}_{\ell}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{N_L(\ell)} \neq h_{N_L(\ell)}\}\right]\right] < \delta_{dec} - \varepsilon_3$$

for some $\varepsilon_3 > 0$, which is sufficient to find h, and therefore $\mathcal{C}_2(h)$, via Lemma 7.8.

As before, this algorithm can be derandomized by going over all random choices which must discover the entire list $\mathcal{L}(g, \mathcal{J}_{q_2}((\delta_0 - \kappa)\delta_2) - \varepsilon)$. ∎

Note that this argument does not place any restriction on δ_0, and in particular, if we choose κ to be much smaller than δ_0, we can decode arbitrarily close to the Johnson radius corresponding to the product bound $\mathcal{J}(\delta_0 \cdot \delta_2)$, even for small values of δ_0 like $1/3$. In contrast, the existing list decoding algorithms for concatenated codes via list recovery of outer Reed-Solomon code [GS02] only approach this Johnson bound when the outer distance is very close to 1.

In fact, we can also decode up to the Johnson bound of product of distances for any outer code that supports list decoding up to Johnson bound via our Covering Lemma/SoS-based machinery. This also includes Tanner code in particular.

9 List Decoding Codes on Square Cayley Complex

First, let us set some notation for the quadripartite left-right square Cayley complex on the group G of size n with generator sets A, B of size d each. Let $\mathcal{S}$ denote the set of squares. The functions X, Y, U, V are 4 bijections from $G \times A \times B$ to $\mathcal{S}$, with the following property:

$$X(g, a, b) = Y(ga, a^{-1}, b) = U(bga, a^{-1}, b) = V(bg, a, b^{-1})$$

The set $X(g, \cdot, b) \subseteq \mathcal{S}$ is defined as $\{X(g, a, b) \ : \ a \in A\}$, and likewise for $X(g, \cdot, \cdot)$, $X(g, a, \cdot)$ and corresponding Y, U, V sets. The sets $X(g, \cdot, \cdot), Y(g, \cdot, \cdot), U(g, \cdot, \cdot)$ and $V(g, \cdot, \cdot)$ should be seen as the analogs of sets $N_L(\ell)$ and $N_R(r)$ from the Tanner codes.

The code on this square Cayley complex is then defined as

$$\begin{aligned}
\mathcal{C}^{SCC} = \{h \in [q]^{\mathcal{S}} \ : \ \forall g \in G, h|_{X(g,\cdot,\cdot)} &\in C_A \otimes C_B, \\
h|_{Y(g,\cdot,\cdot)} &\in C_A \otimes C_B, \\
h|_{U(g,\cdot,\cdot)} &\in C_A \otimes C_B, \\
h|_{V(g,\cdot,\cdot)} &\in C_A \otimes C_B\}
\end{aligned}$$

where C_A and C_B are inner codes of blocklength d each, and $C_A \otimes C_B$ is their tensor code.

For the code defined by left-right Cayley complex with inner codes C_A, C_B with parameters (d, δ_A, r_A) and (d, δ_B, r_B) respectively, the distance of $\mathcal{C}^{SCC}$ is lower bounded [DEL+22] by

$$\delta = \delta_A \delta_B (\max(\delta_A, \delta_B) - \lambda)$$

Note that the distance of tensor code $C_A \otimes C_B$ is at least $\delta_A \cdot \delta_B$.

Theorem 9.1. *For every $\varepsilon > 0$, there is an algorithm based on $\mathcal{O}_{q,d}(1/\varepsilon^4)$ levels of the SoS-hierarchy that runs in time $n^{\mathcal{O}_{q,d}(1/\varepsilon^4)}$ and can list decode $\mathcal{C}^{SCC}$ up to $\mathcal{J}_q(\delta) - \varepsilon$.*

Proof Sketch. We outline the proof by once again focusing on a proof of distance for the appropriate notion of η-good pseudocodewords, and combining this proof with the covering lemma can be done as in the case of Tanner codes. Assume WLOG that $\delta_A \geq \delta_B$.

Consider the SoS relaxation where variables correspond to squares: $\mathbf{Z} = \{Z_{s,j}\}_{s \in \mathcal{S}, j \in [q]}$. The SoS-degree of this relaxation is at least d^2 so that we can enforce all the inner code constraints by making the SoS relaxation respect such constraints explicitly.

Note that for a fixed b, the set of squares $\{X(g, a, b) \ : \ g \in G, a \in A\}$ can be seen as the edges of an expander code, with the inner code as C_A. This motivates the following definition.

A pseudocodeword is called η-good if for all $b \in B$,

$$\mathbb{E}_{g_1, g_2} \left[\widetilde{\mathrm{Cov}} \left[\mathbf{Z}_{X(g_1, \cdot, b)}, \mathbf{Z}_{Y(g_2, \cdot, b)} \right] \right] \leq \eta \tag{13}$$

As we proved before, this implies that for any $h \in \mathcal{C}^{SCC}$,

$$\tau_b = \sqrt{\mathbb{E}_g \left[\widetilde{\mathbb{E}} \left[\mathbf{1}\{\mathbf{Z}_{X(g,\cdot,b)} \neq h_{X(g,\cdot,b)}\} \right] \right]} \sqrt{\mathbb{E}_g \left[\widetilde{\mathbb{E}} \left[\mathbf{1}\{\mathbf{Z}_{Y(g,\cdot,b)} \neq h_{Y(g,\cdot,b)}\} \right] \right]}$$

satisfies

$$\tau_b^2 - (\delta_A - \lambda)\tau_b + \eta \geq 0$$

Let's prove upper and lower bounds on $\mathbb{E}_s[\mathbf{1}\{Z_s \neq h_s\}]$ in terms of τ_b.

$$\mathbb{E}_s\left[\mathbf{1}\{\mathbf{Z}_s \neq h_s\}\right] = \mathop{\mathbb{E}}_{g,a,b}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{X(g,a,b)} \neq h_{X(g,a,b)}\}\right]\right]$$

$$= \mathop{\mathbb{E}}_{g}\left[\widetilde{\mathbb{E}}\left[\Delta(\mathbf{Z}_{X(g,\cdot,\cdot)}, h_{X(g,\cdot,\cdot)})\right]\right]$$

$$\geq \mathop{\mathbb{E}}_{g}\left[\widetilde{\mathbb{E}}\left[0 \cdot \mathbf{1}\{\mathbf{Z}_{X(g,\cdot,\cdot)} = h_{X(g,\cdot,\cdot)}\}\right] + \delta_A \cdot \delta_B \cdot \mathbf{1}\{\mathbf{Z}_{X(g,\cdot,\cdot)} \neq h_{X(g,\cdot,\cdot)}\}\right]$$

$$= \delta_A \delta_B \mathop{\mathbb{E}}_{g}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{X(g,\cdot,\cdot)} \neq h_{X(g,\cdot,\cdot)}\}\right]\right]$$

$$\geq \delta_A \delta_B \mathop{\mathbb{E}}_{g}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{X(g,\cdot,b)} \neq h_{X(g,\cdot,b)}\}\right]\right]$$

where the last inequality is true for any $b \in B$.

Likewise, $\mathbb{E}_s[\mathbf{1}\{\mathbf{Z}_s \neq h_s\}] \geq \delta_A \delta_B \mathbb{E}_g[\widetilde{\mathbb{E}}[\mathbf{1}\{\mathbf{Z}_{Y(g,\cdot,b)} \neq h_{Y(g,\cdot,b)}\}]]$. Therefore,

$$\mathbb{E}_s\left[\mathbf{1}\{\mathbf{Z}_s \neq h_s\}\right] \geq \delta_A \delta_B \sqrt{\mathop{\mathbb{E}}_{g}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{X(g,\cdot,b)} \neq h_{X(g,\cdot,b)}\}\right]\right]} \sqrt{\mathop{\mathbb{E}}_{g}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{Y(g,\cdot,b)} \neq h_{Y(g,\cdot,b)}\}\right]\right]}$$

$$= \delta_A \delta_B \tau_b$$

For the upper bound,

$$\mathbb{E}_s\left[\mathbf{1}\{\mathbf{Z}_s \neq h_s\}\right] = \mathop{\mathbb{E}}_{g,a,b}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{X(g,a,b)} \neq h_{X(g,a,b)}\}\right]\right]$$

$$= \mathop{\mathbb{E}}_{b}\left[\mathop{\mathbb{E}}_{g,a}\left[\widetilde{\mathbb{E}}\left[\mathbf{1}\{\mathbf{Z}_{X(g,a,b)} \neq h_{X(g,a,b)}\}\right]\right]\right]$$

$$\leq \mathop{\mathbb{E}}_{b}\left[\tau_b^2 + \lambda \tau_b + \eta\right]$$

Importantly, the lower bound works for any b, and so if any τ_b is large, we can conclude good distance. Otherwise, all τ_b are small, and then the upper bound works well.

Suppose there is some b for which $\tau_b \geq (\delta_A - \lambda) - \frac{2\eta}{\delta_A - \lambda}$. Then, the distance is at least $\delta_A \delta_B((\delta_A - \lambda) - \frac{2\eta}{\delta_A - \lambda})$.

If not, then all τ_b are at most $\frac{2\eta}{\delta_A - \lambda}$. Then the distance,

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) \leq \mathop{\mathbb{E}}_{b}\left[\tau_b^2 + \lambda \tau_b + \eta\right]$$

$$\leq \mathop{\mathbb{E}}_{b}\left[\frac{4\eta^2}{(\delta_A - \lambda)^2} + \frac{\eta(\delta_A + \lambda)}{\delta_A - \lambda}\right]$$

$$= \frac{4\eta^2}{(\delta_A - \lambda)^2} + \frac{\eta(\delta_A + \lambda)}{\delta_A - \lambda}$$

Putting everything together, if $\lambda \leq \delta_A/3$ and $\eta \leq \delta_A^2/9$, we get that

$$\Delta(\widetilde{\mathbb{E}}[\cdot], h) \geq \delta_A \delta_B(\delta_A - \lambda) - 3\delta_B \eta \geq \delta - 3\eta$$

or

$$\Delta(\widetilde{\mathbb{E}}[\cdot)], h) \leq 3\eta$$

Rest of the decoding argument is as before via covering lemma. We only need to show that the η-good property needed in Eq. (13) can be obtained by conditioning as before. In case we end

up very close to a codeword, we can still unique decode via known unique decoding algorithms [DEL$^+$22] and Lemma 7.4.

The condition in Eq. (13) is that d different average correlations are small. Recall that we ensured in the Tanner code case that covariance is small with probability $1 - \gamma$ where γ can be made arbitrarily small with SoS-degree. By a union bound, we can ensure that the covariance is small across all the d bipartite graphs with probability at least $1 - d \cdot \gamma$. Thus, by paying an additional factor of d in the SoS-degree, we can ensure low covariance across all $b \in B$ as needed.

As before, this argument can be derandomized as well. ∎

Acknowledgements

References

[ABN$^+$92] N. Alon, J. Bruck, J. Naor, M. Naor, and R. Roth. Construction of asymptotically good, low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 28:509–516, 1992. 1, 4, 6

[ADS12] Sanjeev Arora, Constantinos Daskalakis, and David Steurer. Message-passing algorithms and improved lp decoding. *IEEE Transactions on Information Theory*, 58(12):7260–7271, 2012. doi:10.1109/TIT.2012.2208584. 1

[AEL95] N. Alon, J. Edmonds, and M. Luby. Linear time erasure codes with nearly optimal recovery. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 512–519, 1995. doi:10.1109/SFCS.1995.492581. 1, 2, 6, 13

[AJQ$^+$20] Vedat Levi Alev, Fernando Granha Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. List decoding of direct sum codes. In *Proceedings of the 31st ACM-SIAM Symposium on Discrete Algorithms*, pages 1412–1425. SIAM, 2020. 3, 4, 11, 12

[BGG22] Thiago Bergamaschi, Louis Golowich, and Sam Gunn. Approaching the quantum singleton bound with approximate error correction, 2022. arXiv:2212.09935. 3

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science*, pages 472–481, 2011. 12, 23

[BS14] Boaz Barak and David Steurer. Sum-of-squares proofs and the quest toward optimal algorithms. *arXiv preprint arXiv:1404.5236*, 2014. URL: https://arxiv.org/abs/1404.5236. 7

[CCLO23] Xue Chen, Kuan Cheng, Xin Li, and Minghui Ouyang. Improved decoding of expander codes. *IEEE Transactions on Information Theory*, pages 1–1, 2023. doi:10.1109/TIT.2023.3239163. 4

[DEL$^+$22] Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 357–374, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3519935.3520024. 1, 2, 36, 38

[DHK+19]   Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma.
            List decoding with double samplers. In *Proceedings of the 30th ACM-SIAM Symposium
            on Discrete Algorithms*, pages 2134–2153, 2019. 4

[Fel03]   Jon Feldman. *Decoding error-correcting codes via linear programming*. PhD thesis,
            Massachusetts Institute of Technology, 2003. URL: http://hdl.handle.net/1721.
            1/42831. 1

[FKP19]   Noah Fleming, Pravesh Kothari, and Toniann Pitassi. Semialgebraic proofs and efficient
            algorithm design. *Foundations and Trends® in Theoretical Computer Science*, 14(1-2):1–221,
            2019. URL: http://dx.doi.org/10.1561/0400000086, doi:10.1561/0400000086. 1,
            4

[FWK05]   J. Feldman, M.J. Wainwright, and D.R. Karger. Using linear programming to decode
            binary linear codes. *IEEE Transactions on Information Theory*, 51(3):954–972, 2005. doi:
            10.1109/TIT.2004.842696. 1

[Gal62]   R. Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*,
            8(1):21–28, 1962. 1

[GGR09]   P. Gopalan, V. Guruswami, and P. Raghavendra. List decoding tensor products and
            interleaved codes. In *STOC '09*, page 13–22, New York, NY, USA, 2009. Association for
            Computing Machinery. doi:10.1145/1536414.1536419. 4

[GI03]   Venkatesan Guruswami and Piotr Indyk. Linear time encodable and list decodable
            codes. In *Proceedings of the 35th ACM Symposium on Theory of Computing*, 2003. 1

[GI05]   V. Guruswami and P. Indyk. Linear-time encodable/decodable codes with near-optimal
            rate. *IEEE Transactions on Information Theory*, 51(10):3393–3400, 2005. 6, 29

[GKO+17]   Sivakanth Gopi, Swastik Kopparty, Rafael Oliveira, Noga Ron-Zewi, and Shubhangi
            Saraf. Locally testable and locally correctable codes approaching the Gilbert-Varshamov
            bound. In *Proceedings of the 28th ACM-SIAM Symposium on Discrete Algorithms*, SODA
            '17, pages 2073–2091, 2017. 3

[GL18]   Badih Ghazi and Euiwoong Lee. Lp/sdp hierarchy lower bounds for decoding random
            ldpc codes. *IEEE Transactions on Information Theory*, 64(6):4423–4437, 2018. doi:10.
            1109/TIT.2017.2678531. 4

[GR06]   Venkatesan Guruswami and Atri Rudra. Explicit capacity-achieving list-decodable
            codes. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, pages 1–10,
            2006. 3, 32

[GRS23]   Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. Avail-
            able at https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/,
            2023. 1

[GS00]   Venkatesan Guruswami and Madhu Sudan. List decoding algorithms for certain
            concatenated codes. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory
            of Computing*, page 181–190, New York, NY, USA, 2000. Association for Computing
            Machinery. doi:10.1145/335305.335327. 32

[GS01]   Venkatesan Guruswami and Madhu Sudan. Extensions to the johnson bound.
            Manuscript, 2001. 13

[GS02]   V. Guruswami and M. Sudan. Decoding concatenated codes using soft information. In
            Proceedings 17th IEEE Annual Conference on Computational Complexity, pages 148–157,
            2002. doi:10.1109/CCC.2002.1004350. 32, 35

[Gur01]   Venkatesan Guruswami. *List Decoding of Error-Correcting Codes*. PhD thesis, MIT, 2001. 13, 15

[Gur04]   Venkatesan Guruswami. Guest column: Error-correcting codes and expander graphs. *SIGACT News*, 35(3):25–41, September 2004. 1

[Gur06]   Venkatesan Guruswami. Iterative decoding of low-density parity check codes (a survey), 2006. `arXiv:cs/0610022`. 1

[Gur07]   Venkatesan Guruswami. Algorithmic results in list decoding. *Foundations and Trends® in Theoretical Computer Science*, 2(2):107–195, 2007. URL: `http://dx.doi.org/10.1561/0400000007`, `doi:10.1561/0400000007`. 1

[HLW06]   Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, August 2006. 1

[HW18]   Brett Hemenway and Mary Wootters. Linear-time list recovery of high-rate expander codes. *Information and Computation*, 261:202–218, 2018. ICALP 2015. `doi:https://doi.org/10.1016/j.ic.2018.02.004`. 1

[JQST20]   Fernando Granha Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. Unique decoding of explicit ε-balanced codes near the Gilbert–Varshamov bound. In *Proceedings of the 61st IEEE Symposium on Foundations of Computer Science*, 2020. 4, 12

[KMRZS16] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally-correctable and locally-testable codes with sub-polynomial query complexity. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 202–215, New York, NY, USA, 2016. Association for Computing Machinery. `doi:10.1145/2897518.2897523`. 3

[Kop16]   Swastik Kopparty. Distance amplification and list decodable codes. Lecture notes, Spring 2016. URL: `https://sites.math.rutgers.edu/~sk1233/courses/codes-S16/lec7.pdf`. 2

[Lau09]   Monique Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging Applications of Algebraic Geometry (of IMA Volumes in Mathematics and its Applications)*. Springer, 2009. 7

[LPS88]   Alexander Lubotzky, R. Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8:261–277, 1988. 5

[LZ22]   Anthony Leverrier and Gillés Zémor. Quantum tanner codes. *arXiv preprint arXiv:2202.13641*, 2022. 1

[MRRZ⁺20] J. Mosheiff, N. Resch, N. Ron-Zewi, S. Silas, and M. Wootters. Ldpc codes achieve list decoding capacity. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 458–469, Los Alamitos, CA, USA, nov 2020. IEEE Computer Society. `doi:10.1109/FOCS46700.2020.00050`. 1

[O'D17]   Ryan O'Donnell. SOS Is Not Obviously Automatizable, Even Approximately. In Christos H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 59:1–59:10, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. URL: `http://drops.dagstuhl.de/opus/volltexte/2017/8198`, `doi:10.4230/LIPIcs.ITCS.2017.59`. 7

[PK22]   Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In *Proceedings of the 54th ACM Symposium on Theory of Computing*, pages 375–388, 2022. 1

[RR22]   S. Richelson and S. Roy. List-decoding xor codes near the johnson bound. Technical Report TR22-069, Electronic Colloquium on Computational Complexity, 2022. URL: https://eccc.weizmann.ac.il/report/2022/069. 4, 12

[RU08]   Tom Richardson and Rüdiger Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008. doi:10.1017/CBO9780511791338. 1

[RW17]   Prasad Raghavendra and Benjamin Weitz. On the bit complexity of sum-of-squares proofs. In *Proceedings of the 44th International Colloquium on Automata, Languages and Programming*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. 7

[RZWZ21]   Noga Ron-Zewi, Mary Wootters, and Gillés Zémor. Linear-time erasure list-decoding of expander codes. *IEEE Transactions on Information Theory*, 67(9):5827–5839, 2021. doi:10.1109/TIT.2021.3086805. 1

[SR03]   V. Skachek and R.M. Roth. Generalized minimum distance iterative decoding of expander codes. In *Proceedings 2003 IEEE Information Theory Workshop (Cat. No.03EX674)*, pages 245–248, 2003. doi:10.1109/ITW.2003.1216740. 6

[SS96]   M. Sipser and D. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6):1710–1722, 1996. Preliminary version in *Proc. of FOCS'94*. 1, 4, 5, 6

[Tan81]   R. Tanner. A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5):533–547, 1981. doi:10.1109/TIT.1981.1056404. 1, 5

[TS17]   Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, STOC 2017, pages 238–251, New York, NY, USA, 2017. ACM. 1, 12

[Vad12]   Salil P. Vadhan. *Pseudorandomness*. Now Publishers Inc., 2012. 1

[Zém01]   G. Zémor. On expander codes. *IEEE Transactions on Information Theory*, 47(2):835–837, 2001. doi:10.1109/18.910593. 1, 2, 5, 6