

Sparser Abelian High Dimensional Expanders

Yotam Dikstein 

Institute for Advanced Study, Princeton, NJ, USA

Siqi Liu 

Institute for Advanced Study, Princeton, NJ, USA

Avi Wigderson 

Institute for Advanced Study, Princeton, NJ, USA

Abstract

The focus of this paper is the development of new elementary techniques for the construction and analysis of high dimensional expanders. Specifically, we present two new explicit constructions of Cayley high dimensional expanders (HDXs) over the abelian group $\mathbb{F}_2^n$. Our expansion proofs use only linear algebra and combinatorial arguments.

The first construction gives local spectral HDXs of any constant dimension and subpolynomial degree $\exp(n^\varepsilon)$ for every $\varepsilon > 0$, improving on a construction by Golowich [50] which achieves $\varepsilon = 1/2$. [50] derives these HDXs by sparsifying the *complete* Grassmann poset of subspaces. The novelty in our construction is the ability to sparsify *any* expanding Grassmann posets, leading to iterated sparsification and much smaller degrees. The sparse Grassmannian (which is of independent interest in the theory of HDXs) serves as the generating set of the Cayley graph.

Our second construction gives a 2-dimensional HDX of any polynomial degree $\exp(\varepsilon n)$ for any constant $\varepsilon > 0$, which is simultaneously a spectral expander and a coboundary expander. To the best of our knowledge, this is the first such non-trivial construction. We name it the Johnson complex, as it is derived from the classical Johnson scheme, whose vertices serve as the generating set of this Cayley graph. This construction may be viewed as a derandomization of the recent random geometric complexes of [74]. Establishing coboundary expansion through Gromov’s “cone method” and the associated isoperimetric inequalities is the most intricate aspect of this construction.

While these two constructions are quite different, we show that they both share a common structure, resembling the intersection patterns of vectors in the Hadamard code. We propose a general framework of such “Hadamard-like” constructions in the hope that it will yield new HDXs.

2012 ACM Subject Classification Theory of computation; Mathematics of computing → Spectra of graphs

Keywords and phrases Local spectral expander, coboundary expander, Grassmannian expander

Digital Object Identifier 10.4230/LIPIcs.CCC.2025.7

Related Version *Full Version*: <https://arxiv.org/abs/2411.08839>

Funding *Yotam Dikstein*: This material is based upon work supported by the National Science Foundation under Grant No. DMS-1926686.

Siqi Liu: Supported by Minerva Research Foundation Member Fund. This work was partially conducted at the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) with support from DIMACS and Rutgers University.

Avi Wigderson: Supported by NSF grant CCF-1900460.

Acknowledgements We thank Louis Golowich for helpful discussions and Gil Melnik for assistance with the figures.

1 Introduction

Expander graphs are of central importance in many diverse parts of computer science and mathematics. Their structure and applications have been well studied for half a century, resulting in a rich theory (see e.g. [56, 76]). First, different definitions of expansion (spectral, combinatorial, probabilistic) are all known to be essentially equivalent. Second, we know that



© Yotam Dikstein, Siqi Liu, and Avi Wigderson;
licensed under Creative Commons License CC-BY 4.0

40th Computational Complexity Conference (CCC 2025).

Editor: Srikanth Srinivasan; Article No. 7; pp. 7:1–7:98

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



expanders abound; a random sparse graph is almost surely an expander [89]. Finally, there is a variety of methods for constructing and analyzing explicit expander graphs. Initially, these relied on deep algebraic and number theoretic methods [84, 45, 78]. However, the quest for elementary (e.g. combinatorial and linear-algebraic) constructions has been extremely fertile, and resulted in major breakthroughs in computer science and math; the *zigzag* method of [93] lead to [33, 92, 96], and the *lifting* method of [14] lead to [83, 82].

In contrast, the expansion of sparse high dimensional complexes (or simply hypergraphs) is still a young and growing research topic. The importance of high dimensional expanders (HDXs) in computer science and mathematics is consistently unfolding, with a number of recent breakthroughs which essentially depend on them, such as locally testable codes [34, 88], quantum coding [6] and Markov chains [5] as a partial list. High dimensional expanders differ from expander graphs in several important ways. First, different definitions of expansion are known to be non-equivalent [53]. Second, they seem rare, and no natural distributions of complexes are known which almost surely produce bounded-degree expanding complexes. Finally, so far the few existing explicit constructions of bounded degree HDXs use algebraic methods [12, 16, 72, 79, 68, 86, 25]. No elementary constructions are known where high dimensional expansion follows from a clear (combinatorial or linear algebraic) argument. This demands further understanding of HDXs and motivates much prior work as well as this one.

In this work we take a step towards answering this question by constructing sparser HDXs by elementary means. We construct sparse HDXs with two central notions of high dimensional expansion: local spectral expansion and coboundary expansion.

To better discuss our results, we define some notions of simplicial complexes and expansion. For simplicity, we only define them in the 2-dimensional case, though many of our results extend to arbitrary constant dimensions. For the full definitions see Section 2. A 2-dimensional simplicial complex is a hypergraph $X = (V, E, T)$ with vertices V , edges E and triangles T (sets of size 3). We require that if $t \in T$ then all its edges are in E , and all its vertices are in V . We denote by N the number of vertices. Given a vertex $u \in V$, its degree is the number of edges adjacent to it. An essential notion for HDXs is the *local* property of its links. A *link* of a vertex u is the graph $G_u = (V_u, E_u)$ whose set of vertices is the neighborhood of u ,

$$V_u = \{v \in V \mid \{u, v\} \in E\}.$$

The edges correspond to the triangles incident on u :

$$E_u = \{\{v, w\} \mid \{u, v, w\} \in T\}.$$

In this paper we focus on a particularly simple class of complexes, called Cayley complexes over $\mathbb{F}_2^n$. These are simplicial complexes whose underlying graph (V, E) is a Cayley graph over $\mathbb{F}_2^n$, and so have $N = 2^n$ vertices. It is well-known that constant-degree Cayley complexes over abelian groups cannot be high dimensional expanders. In fact, logarithmic degree is needed for a 1-dimensional complex, namely a graph, to be an expander [3]. In particular, random Cayley graphs over $\mathbb{F}_2^n$ with degree $\Theta(n)$ (logarithmic in the number of vertices) are expanders with high probability. However, no abelian Cayley complex constructions with remotely close degree are known to have higher dimensional expansion, and a guiding question of this paper is to develop techniques for constructing and analyzing abelian Cayley HDXs of smallest possible degree.

In the rest of the introduction we will explain our constructions. In Section 1.1 we present our construction of local spectral Cayley complexes with arbitrarily good spectral expansion and a subpolynomial degree. In Section 1.2 we present our family of Cayley complexes that are both coboundary expanders and non-trivial local spectral expanders,

with arbitrarily small polynomial degree. In Section 1.3 we discuss the common structure of our two constructions that relies on the Hadamard code. In Section 1.4 we present some open questions.

1.1 Local spectral expanders

We start with definitions and motivation, review past results and then state ours. The first notion of high dimensional expansion we consider is *local spectral expansion*. The spectral expansion of a graph G is the second largest eigenvalue of its random walk matrix in absolute value, denoted $\lambda(G)$. Local spectral expansion of complexes is the spectral expansion of every link.

► **Definition 1** (Local spectral expansion). *Let $\lambda \geq 0$. A 2-dimensional simplicial complex $X = (V, E, T)$ is a λ -local spectral expander if the underlying graph (V, E) is connected and for every $u \in V$, $\lambda(X_u) \leq \lambda$.*

Local spectral expansion is the most well-known definition of high dimensional expansion. It was first introduced implicitly by [47] for studying the real cohomology of simplicial complexes. A few decades later, a few other works observed that properties of specific local spectral expanders are of interest to the computer science community [79, 60, 62]. The definition above was given in [38], for the purpose of constructing agreement tests. These are strong property tests that are a staple in PCPs (see further discussion in Section 1.2). Since then they have found many more applications that we now discuss.

Applications of local spectral expanders

One important application of local spectral expansion is in approximate sampling and Glauber dynamics. Local spectral expansion implies optimal spectral bounds on random walks defined on the sets of the complex. These are known as the Glauber dynamics or the up-down walks. This property has lead to a lot of results in approximate sampling and counting. These include works on Matroid sampling [5], random graph coloring [20] and mixing of Glauber dynamics on the Ising model [4].

Local spectral expanders have also found applications in error correcting codes. Their sampling properties give rise to construction of good codes with efficient list-decoding algorithms [1, 37, 59]. One can also use them to construct locally testable codes that sparsify the Reed-Muller codes [39].

The list decoding algorithms in [1, 59] rely on their work on constraint satisfaction problems on local spectral expanders. These works generalize classical algorithms for solving CSPs on dense hypergraphs, to algorithms that solve CSPs on local spectral expanders. These works build on the “dense-like” properties of these complexes, and prove that CSPs on local spectral expanders are *easy*. It is interesting to note that local spectral expanders have rich enough structure so that one can also construct CSPs over them that are also *hard* for Sum-of-Squares semidefinite programs [35, 57]¹.

Some of the most important applications of local spectral expansion are constructions of derandomized agreement tests [38, 26, 51, 9, 31] and even PCPs [11]. Most of these works also involve coboundary expansion so we discuss these in Section 1.2.

¹ In fact, [1] uses Sum-of-Squares to solve CSPs on local spectral expanders. It seems contradictory, but the hardness results in [35, 57] use CSPs where the variables are the edges of the HDX. The CSPs in [1] use its vertices as variables.

Local spectral expanders have other applications in combinatorics as well. They are especially useful in sparsifying central objects like the complete uniform hypergraph and the boolean hypercube. The HDX based sparsifications maintain many of their important properties such as spectral gap, Chernoff bounds and hypercontractivity (see [62, 30, 67, 54, 8, 32] as a partial list).

Previous constructions

Attempts to construct bounded degree local spectral expanders with sufficient expansion for applications have limited success besides the algebraic constructions mentioned above [12, 16, 72, 79, 68, 86, 25]. Random complexes in the high dimensional Erdős–Rényi model defined by [73] are *not* local spectral expanders with overwhelming probability when the average degree is below $N^{1/2}$ and no other bounded degree model is known. The current state of the art is the random geometric graphs on the sphere [74], that achieve non-trivial local spectral expansion and arbitrarily small polynomial degree (but their expansion is also bounded from below, see discussion in Section 1.2). Even allowing an unbounded degree, the only other known construction that is non-trivially sparse complexes is the one in [50] mentioned in the abstract.

Our result

The first construction we present is a family of Cayley local spectral expanders. Namely,

► **Theorem 2.** *For every $\lambda, \varepsilon > 0$ and integer $d \geq 2$ there exists an infinite family of d -dimensional Cayley λ -local spectral expanders over the vertices $\mathbb{F}_2^n$ with degree at most 2^{n^ε} .*

This construction builds on an ingenious construction by Golowich [50], which proved this theorem for $\varepsilon = \frac{1}{2}$. As far as we know, this is the sparsest construction that achieves an arbitrarily small λ -local spectral expansion other than the algebraic constructions.

Proof overview

Our construction is based on the well studied Grassmann poset, the partially ordered set of subspaces of $\mathbb{F}_2^n$, ordered by containment. This object is the vector space equivalent of high dimensional expander, and was previously studied in [30, 69, 46, 50]. To understand our construction, we first describe the one in [50]. Golowich begins by sparsifying the Grassmann poset, to obtain a subposet. The generators of the Cayley complex in [50] are all (non-zero vectors in) the one-dimensional subspaces in the subposet. The analysis of expansion in [50] depends on the expansion properties of the subposet. The degree is just the number of one-dimensional vector spaces in the subposet, which is $2^{n^{1/2}}$.

Our construction takes a modular approach to this idea. We modify the poset sparsification in [50] so that instead of the entire Grassmann poset, we can plug in *any* subposet Y of the Grassmann, and obtain a sparsification Y' whose size depends on Y (not the complete Grassmann).

Having this flexible sparsification step, we *iterate*. We start with Y_0 , the complete Grassmann poset, we obtain a sequence $Y_1, Y_2, \dots, Y_m$ of sparser and sparser subposets. The vectors in Y_i are low-rank matrices whose rows and columns are vectors in subspaces of Y_{i+1} .

We comment that while this is a powerful paradigm, it does have a drawback. The dimension of Y_i is logarithmic in the dimension of Y_{i-1} (the dimension of Y is the maximal dimension of a space inside Y). Thus for a given complex Y_0 we can only perform this

sparsification a constant number of steps keeping ε constant. Our infinite family is generated by taking an infinite family of Y_0 's, and using our sparsification procedure on every one of them separately.

This construction of sparsified posets produces local spectral expanders of any constant dimension, as observed by [50]. The higher dimensional sets in the resulting simplicial complex correspond to the higher dimensional subspaces of the sparsification we obtain. We refer the reader to Section 3 for more details.

The analysis of the local spectral expansion is delicate, so we will not describe it here. One property that is necessary for it is the fact that every rank- r matrix has many decompositions into sums of rank-1 matrices. Therefore, to upper bound spectral expansion we study graphs that arise from decompositions of matrices.

In addition, we wish to highlight a local-to-global argument in our analysis that uses a theorem by [81]. Given a graph $G = (V, E)$ that we wish to analyze, we find a set of expanding subgraphs $\{H_1, H_2, \dots, H_m\}$ of G (that are allowed to overlap). We consider the *decomposition* graph whose vertices are V and whose edges are all $\{v, v'\}$ such that there exists an H_i such that $v, v' \in H_i$. If every H_i is an expander and the decomposition graph is also an expander, then G itself is an expander [81]. This decomposition is particularly useful in our setup. The graphs we need to analyze in order to prove the expansion of our complexes, can be decomposed into smaller subgraphs H_i . These H_i 's are isomorphic to graphs coming from the original construction of [50]. Using this decomposition we are able to reduce the expansion of the links to the expansion of some well studied containment graphs in the original Grassmann we sparsify.

► **Remark 3 (Degree lower bounds for Cayley local spectral expanders).** The best known lower bound on the degree of Cayley λ -local spectral expanders over $\mathbb{F}_2^n$ is $\Omega\left(\frac{n}{\lambda^2 \log(1/\lambda)}\right)$ [2]. This bound is obtained simply because the underlying graph of a Cayley λ -local spectral expander is an $O(\lambda)$ -spectral expander itself [87]. In Section I we investigate this question further, and provide some additional lower bounds on the degree of Cayley complexes based on their link structure. We prove that if the Cayley complex has a connected link then its degree is lower bounded by $2n - 1$ (compared to n if the link is not connected). If the link mildly expands, we provide a bound that grows with the degree of a vertex *inside the link*. However, these bounds do not rule out Cayley λ -local spectral expanders whose generator set has size $O\left(\frac{n}{\lambda^2 \log(1/\lambda)}\right)$. The best trade-off between local spectral expansion and the degree of the complex is still open.

1.2 Our coboundary expanders

Our next main result is an explicit construction of a family of 2-dimensional complexes with an arbitrarily small polynomial degree, that have both non-trivial local spectral expansion and coboundary expansion. To the best of our knowledge, this is the first such nontrivial construction. Again, these are Cayley complexes of $\mathbb{F}_2^n$.

Coboundary expansion is a less understood notion compared to local spectral expansion. Therefore, before presenting our result we build intuition for it slowly by introducing a testing-based definition of coboundary expanders. After the definition, we will discuss the motivation behind coboundary expansion and applications of known coboundary expanders.

As mentioned before, there are several nonequivalent definitions of expansion for higher-dimensional complexes. In particular, coboundary expansion generalizes the notion of edge expansion in graphs to simplicial complexes. Coboundary expansion was defined independently by [73],[85] and [52]. The original motivation for this definition came from

discrete topology; more connections to property testing were discovered later. We will expand on these connections after giving the definition. This definition we give is equivalent to the standard definition, but is described in the language of property testing rather than cohomology.

The more general and conventional definition for arbitrary groups Γ can be found in Section 5.1.

► **Definition 4.** Let $X = (V, E, T)$ be a 2-dimensional simplicial complex. Consider the code $B^1 \subseteq \mathbb{F}_2^E$ whose generating matrix is the vertex-edge incidence matrix $A_X \in \mathbb{F}_2^{E \times V}$, and the local test $\mathcal{T}$ of membership in this code for B^1 given by first sampling a uniformly random triangle $\{u, v, w\} \in T$ and then accepting $f \in \mathbb{F}_2^E$ if and only if $f(\{u, v\}) + f(\{v, w\}) + f(\{u, w\}) = 0$ over $\mathbb{F}_2$.

Let $\beta > 0$. X is a β -coboundary expander over $\mathbb{F}_2$ if

$$\forall f \in \mathbb{F}_2^E, \mathbb{P}[\mathcal{T} \text{ rejects } f] \geq \beta \cdot \text{dist}(f, B^1),$$

where $\text{dist}(f, B^1) = \min_{c \in B^1} \mathbb{P}_{e \sim \text{Unif}(E)}[f(e) \neq c(e)]$ is the relative distance from f to the closest codeword in B^1 .

Observe that the triangles are parity checks of B^1 , and so if the input f is a codeword, the test always accepts. A coboundary expander's local test should reject inputs f which are “far” from the code with probability proportional to their distance from it. The proportionality constant β captures the quality of the tester (and coboundary expansion). We note that although the triangles are *some* parity checks of the code B^1 , they do not necessarily span *all* parity checks. In such cases, X is not a coboundary expander for any $\beta > 0$.

► **Definition 5.** A set of 2-dimensional simplicial complexes $\{X_n\}$ is a family of coboundary expanders if there exists some $\beta > 0$ such that for all n , X_n is a β -coboundary expander.

Applications of coboundary expansion

We mention three different motivations to the study of coboundary expansion: from agreement testing, discrete geometry, and algebraic topology.

Coboundary expanders are useful for constructing derandomized agreement tests for the low-soundness (list-decoding) regime. As we mentioned before, agreement tests (also known as derandomized direct product tests) are a strong property test that arise naturally in many PCP constructions [33, 90, 48, 58] and low degree tests [94, 7, 91]. They were introduced by [48]. In these tests one gets oracle access to “local” partial functions on small subsets of a variable set, and is supposed to determine by as few queries as possible, whether these functions are correlated with a “global” function on the whole set of variables. The “small soundness regime”, is the regime of tests where one wants to argue about closeness to a “global” function even when the “local” partial functions pass the test with very small success probability (see [36] for a more formal definition). Works such as [36, 58, 40] studied this regime over complete simplicial complexes and Grassmanians, but until recently these were essentially the only known examples. Works by [51, 10, 27] reduced the agreement testing problems to unique games constraint satisfaction problems, and showed existence of a solution via coboundary expansion (over some symmetric groups), following an idea by [41]. This lead to derandomized tests that come from bounded degree complexes [31, 9].

In discrete geometry, a classical result asserts that for any n points in the plane, there exists a point that is contained in at least a constant fraction of the $\binom{n}{3}$ triangles spanned by the n points [13]². In the language of complexes, for every embedding of the vertices of

² This can also be generalized to any dimension by replacing triangles with simplices [15].

the *complete* 2-dimensional complex to the plane, such a *heavily covered point* exists. One can ask whether such a point exists even when one allows the edges in the embedding to be arbitrary continuous curves. If for every embedding of a complex X to the plane (where the edges are continuous curves), there exists a point that lays inside a constant fraction of the triangles of X , we say X has the *topological overlapping property*. The celebrated result by Gromov states that the complete complex indeed has this property [52]. It is more challenging to prove this property for a given complex.

In [52], Gromov asked whether there exists a family of bounded-degree complexes with the topological overlapping property. Towards finding an answer, Gromov proved that *every* coboundary expander has this property. This question has been extremely difficult. An important progress is made in [43], where they defined *cosystolic expansion*, a relaxation of coboundary expansion, and proved that this weaker expansion also implies the topological overlapping property. The problem was eventually resolved in [60] for dimension 2 and in [44] for dimension > 2 where the authors show that the [79] Ramanujan complexes are bounded-degree cosystolic expanders.

Coboundary expansion has other applications in algebraic topology as well. Linial, Meshulam, and Wallach defined coboundary expansion independently, initiating a study on high dimensional connectivity and cohomologies of random complexes [73, 85]. Lower bounding coboundary expansion turned out to be a powerful method to argue about the vanishing of cohomology in high dimensional random complexes [42, 55].

Known coboundary expanders

Most of the applications mentioned above call for complexes that are simultaneously coboundary expanders and non-trivial local spectral expanders. So far there are no known constructions of such complexes with arbitrarily small polynomial degree even in dimension 2³. Here we summarize some known results.

Spherical buildings of type A_d are dense complexes that appear as the links of the Ramanujan complexes in [79]. The vertex set of a spherical building consists of all proper subspaces of some ambient vector space $\mathbb{F}_q^d$. [52] proved that spherical buildings are coboundary expanders (see also [77]). Geometric lattices are simplicial complexes that generalize spherical buildings. They were first defined in [71]. Most recently, [28] show that geometric lattices are coboundary expanders. [68] show that the 2-dimensional vertex links of the coset geometry complexes from [66] are coboundary expanders.

If we restrict our interest to coboundary expanders over a single finite group instead of all groups, bounded degree coboundary expanders that are local spectral expanders are known. [60] solved this for the $\mathbb{F}_2$ case conditioned on a conjecture by Serre. [19] constructed such complexes for the $\mathbb{F}_2$ unconditionally, and their idea was extended by [31, 9] to every *fixed* finite group, which lead to the aforementioned agreement soundness result. Unfortunately, their approach cannot be leveraged to constructing a coboundary expander with respect to all groups simultaneously.

We note that the coboundary expanders mentioned above are also local spectral expanders and that if we do not enforce the local spectral expansion property, coboundary expanders are trivial yet less useful.

³ By coboundary expanders, we are referring to complexes that have coboundary expansion over *every* group, not just $\mathbb{F}_2$.

To the best of our knowledge, all known constructions of 2-dimensional coboundary expanders (over all groups) that are also non-trivial local spectral expanders have vertex degree at least N^c for some fixed $c > 0$, where N is the number of vertices (see e.g. [79, 66]).⁴ The diameter of all of those complexes is at most a fixed constant, which implies this lower bound on the maximal degree. In this work, we give the first such construction with arbitrarily small polynomial degree.

Our result

We now state our main result in this subsection. For every integer $k > 0$, and constant $\varepsilon \in (0, \frac{1}{2}]$ we construct a family of k -dimensional simplicial complexes $\{X_{\varepsilon,n}\}_n$ called the Johnson complexes. For any k -dimensional simplicial complex X , we use the notation $(X)^{\leq 2}$ (the 2-skeleton of X) to denote the 2-dimensional simplicial complex consists of the vertex, edge, and triangle sets of X .

► **Theorem 6** (Informal version of Theorem 86, Lemma 59, and Lemma 91). *For any integer $k \geq 2$, $\varepsilon \in (0, \frac{1}{2}]$, the 2-skeletons of the Johnson complexes $X_{\varepsilon,n}$ are $(\frac{1}{2} - \frac{\varepsilon}{2})$ -local spectral expanders and $\Omega(\varepsilon)$ -coboundary expanders (over every group Γ).*

Moreover, if $k \geq 3$, the 2-skeletons of $X_{\varepsilon,n}$'s vertex links are $\frac{1}{85}$ -coboundary expanders.

Fix $\varepsilon \in (0, \frac{1}{2}]$. For every n satisfying $4 \mid \varepsilon n$, we briefly describe the 2-skeletons of $X_{\varepsilon,n}$. The underlying graph of $X_{\varepsilon,n}$ is simply the Cayley graph $\text{Cay}(H; S_\varepsilon)$ where $H \subset \mathbb{F}_2^n$ consists of vectors with even Hamming weight, and S_ε is the set of vectors with Hamming weight εn . Thus the number of vertices in the graph is 2^{n-1} while the vertex degree is $\binom{n}{\varepsilon n} < 2^{nh(\varepsilon)}$ ⁵. The triangles are given by:

$$X_{\varepsilon,n}(2) = \{\{x, x + s_1, x + s_2\} \mid x \in \mathbb{F}_2^n, s_1, s_2, s_1 + s_2 \in S_\varepsilon\}.$$

Observe that the link of every vertex is isomorphic to the classically studied Johnson graph $J(n, m, m/2)$ of m -subsets of $[n]$, that are connected if their intersection is $m/2$ (for $m = \varepsilon n$). We will use some of the known properties of this graph in our analysis.

We additionally show that when $2^k \mid \varepsilon n$, we can extend the above construction to get k -dimensional simplicial complexes $X_{\varepsilon,n}$ with the exact same vertex set, edge set, and triangle set as defined above. Moreover we show that for every integer $0 \leq m \leq k - 2$, the link of an m -face in the complex is also a non-trivial local spectral expander (Lemma 59).

► **Remark 7.** Let $N = 2^{n-1}$. We note that all vertices in $X_{\varepsilon,n}$ are in $\text{poly}(N)$ edges and all vertices in the links of $X_{\varepsilon,n}$ are in $\text{poly}(N)$ edges. We remark that using a graph sparsification result due to [21], we can randomly subsample the triangles in $X_{\varepsilon,n}$ to obtain a random subcomplex which is still a local spectral expander with high probability but whose links have vertex degree $\text{polylog}(N)$. A detailed discussion can be found in Section F.

Before giving a high-level overview of the proof for Theorem 6, we describe a general approach for showing coboundary expansion called the *cone method*. Appearing implicitly in [52], it was used in a variety of works (see [77, 71, 68] as a partial list). We also take this approach to show that the Johnson complexes are coboundary expanders.

⁴ To the best of our knowledge, this constant is $c = 0.5$.

⁵ Here h is the binary entropy function.

Coboundary expansion for links of the [50] Cayley local spectral expanders

Even though most of our results on coboundary expansion are for the Johnson complexes, we can also use the methods in this paper to prove that the vertex links of the [50] Cayley local spectral expanders are coboundary expanders. This implies that these Cayley local spectral expanders are cosystolic expanders (see [60, 28]), the relaxation of coboundary expansion mentioned above. This makes them useful for constructing agreement tests and complexes with the topological overlapping property. We note that [50] observed that these complexes are not coboundary expanders so this relaxation is necessary. A definition of cosystolic expansion and a detailed proof can be found in Section 6.3.

Cones and isoperimetry

Recall first the definition of β -coboundary expansion above. We start with the code B^1 of edge functions f which arise from vertex labelings by elements of some group Γ . This implies that composing (in order) the values of $f \in B^1$ along the edges of any cycle will give the identity element $id \in \Gamma$. This holds in particular for triangles, which are our “parity checks”.

We digress to discuss an analogous situation in geometric group theory, of the *word problem* for finitely presented groups. In this context, we are given a word in the generators of a group and need to check if it simplifies to the identity under the given relations. In our context the given word is the labeling of f along some cycle, and a sequence of relations (here, triangles) is applied to successively “contract” this word to the identity. The tight upper bound on the number of such contractions in terms of the length of the word (called the *Dehn function* of the presentation), captures important properties of the group, e.g. whether or not it is hyperbolic. This ratio between the contraction size and the word length is key to the cone method.

One convenient way of capturing the number of contractions is the so-called “van Kampen diagram”, which simply “tiles” the cycle with triangles in the plane (all edges are consistently labeled by group elements). The van Kampen lemma ensures that if a given word can be tiled with t tiles, then there is a sequence of t contractions that reduce it to the identity [97]. The value in this viewpoint is that tiling is a static object, which can be drawn on the plane, and allows one to forget about the ordering of the contractions. We will make use of this in our arguments, and for completeness derive the van Kampen lemma in our context. Note that a bound on the minimum number of tiles (a notion of area) needed to cover any cycle of a given length (a notion of boundary length) can be easily seen as an *isoperimetric inequality*! The cone method will require such isoperimetric inequality, and the Dehn function gives an upper bound on it.

Back to β -coboundary expansion. Here the given f may not be in B^1 , and we need to prove that if it is “far” from B^1 , then the proportion of triangles which do not compose to the identity on f will be at least $\beta \text{dist}(f, B^1)$. The cone method localizes this task. To use this method, one needs to specify a family of cycles (also called a cone) in the underlying graph of the complex. Gromov observed that the complex’s coboundary expansion has a lower bound that is inverse proportional to the number of triangles (in the complex) needed to tile a cycle from this cone [52]. This number is also referred to as the cone area. Thus, bounding the coboundary expansion of the complex reduces to computing the cone area of some cone. Needless to say, an upper bound on the Dehn function, which gives the worst case area for *all* cycles, certainly suffices for bounding the cone area of any cone.⁶

⁶ We note that in a sense the converse is also true - computing the cone area for cones with cycles of

Proof overview

It remains to construct a cone in $X_{\varepsilon,n}$ and upper bound its cone area. We now provide a high level intuition for our approach. First observe that the diameter of the complex $X = X_{\varepsilon,n}$ is $\Theta(1/\varepsilon)$. The proof then proceeds as follows.

1. We move to a denser 2-dimensional complex X' whose underlying graph is the Cayley graph $\text{Cay}(H; S_{\leq \varepsilon})$ where $H \subset \mathbb{F}_2^n$ consists of vectors with even Hamming weight, and $S_{\leq \varepsilon}$ is the set of vectors with even Hamming weight $\leq \varepsilon n$. The triangle set consists of all 3-cliques in the Cayley graph. We note that X' has the same vertex set as X but has more edges and triangles than the Johnson complex.
2. Then we carefully construct a cone $\mathcal{A}'$ in X' with cone area $\Theta(1/\varepsilon)$.
3. We show that every edge in X' translates to a length-2 path in X , and every 3-cycle (boundary of a triangle) in X' translates to a 6-cycle in X which can be tiled by $O(1)$ triangles from X .
4. We translate the cone $\mathcal{A}'$ in X' into a cone $\mathcal{A}$ in X by replacing every edge in the cycles of $\mathcal{A}'$ with a certain length-2 path in X . Thus every cycle $C \in \mathcal{A}$ can be tiled by first tiling its corresponding cycle $C' \in \mathcal{A}'$ with $O(1/\varepsilon)$ X' triangles and then tiling each of the X' triangles with $O(1)$ X triangles. Thereby we conclude that $\mathcal{A}$ has cone area $\Theta(1/\varepsilon)$.

Local spectral expansion and comparison to random geometric complexes

These Johnson complexes also have non-trivial local spectral expansion. While they do not achieve arbitrarily small λ -local spectral expansion, they do pass the $\lambda < \frac{1}{2}$ barrier. From a combinatorial point of view, $\lambda < \frac{1}{2}$ is an important threshold. For $\lambda \geq \frac{1}{2}$, there are elementary bounded-degree constructions of λ -local spectral expanders [22, 17, 23, 75, 49] but these fail to satisfy any of the local-to-global properties in HDXs. For $\lambda < \frac{1}{2}$ all these constructions break. For this regime, all bounded degree constructions rely on algebraic tools. This is not by accident; below $\frac{1}{2}$ there are a number of high dimensional global properties suddenly begin to hold. For instance, a theorem by [47] implies that when $\lambda < \frac{1}{2}$ all real cohomologies of the complex vanish. Another example is the “Trickle-Down” theorem by [87] which says that the expansion of the skeleton (V, E) is non-trivial whenever $\lambda < \frac{1}{2}$. These are strong properties, and we do not know how to produce them in elementary constructions. Therefore, when constructions of local spectral expanders are considered non-trivial only when $\lambda < \frac{1}{2}$.

To show local spectral expansion, we first note that the Johnson complex is symmetric for each vertex. Hence, it suffices to focus on the link of $0 \in \mathbb{F}_2^n$. Because the triangle set T is well-structured, we can show that the link graph of 0 is isomorphic to tensor products of Johnson graphs (Proposition 62). This allows us to use the theory of association schemes to bound their eigenvalues.

We also note that since all boolean vectors in $\{0, 1\}^n$ lie on a sphere in $\mathbb{R}^n$ centered at $[\frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}]$, the Johnson complex $X_{\varepsilon,n}$ can be viewed as a geometric complex whose vertices are associated with points over a sphere and two vertices form an edge if and only if their corresponding points' L_2 distance is $\sqrt{\varepsilon n}$. Previously [74] prove that randomized 2-dimensional geometric complexes are local spectral expanders. Comparing the two constructions, we conclude that Johnson complexes are sparser local spectral expanders than random geometric complexes.

“minimal” length suffices for computing the Dehn function. We indeed give such a strong bound.

1.3 The common structure between the two constructions

Taking a step back, we wish to highlight that the two constructions in our paper share a common structure. Both of these constructions can be described via “induced Hadamard encoding”. For a (not necessarily linear) function $s : \mathbb{F}_2^d \rightarrow \mathbb{F}_2^n$, the “induced” Hadamard encoding is the linear map $\widehat{s} : \mathbb{F}_2^d \rightarrow \mathbb{F}_2^n$ given by

$$\widehat{s}(x) := \sum_{v \in \mathbb{F}_2^d} \langle x, v \rangle s(v),$$

where $\langle x, v \rangle = \sum_{i=1}^d x_i v_i$ and the sum is over $\mathbb{F}_2$. Our two constructions of Cayley HDXs take the generating set of the Cayley graph to be all bases of spaces $\text{Im}(\widehat{s}) \subseteq \mathbb{F}_2^n$, for carefully chosen functions s as above; the choice determines the construction. We note that the “orthogonality” of vectors of the Hadamard code manifests itself very differently in the two constructions: in the Johnson complexes it can be viewed via the Hamming weight metric, while in the other construction it may be viewed via the matrix rank metric. We connect the structure of links in our constructions, to the restrictions of these induced Hadamard encodings to affine subspaces of $\mathbb{F}_2^n$. Using this connection we show that one can decompose the link into a tensor product of simpler graphs, which are amenable to our analysis. A special case of this observation was also used in the analysis of the complexes constructed by [50].

1.4 Open questions

Local spectral expanders

As mentioned above, we do not know how sparse a Cayley local spectral expander can be. To what extent can we close the gaps between the lower and upper bounds for various types of abelian Cayley HDXs? In particular, can we show that any nontrivial abelian Cayley expanders must have degree $\omega(n)$? Conversely, can we construct Cayley HDXs where the degree is upper bounded by some polynomial in n ? Limits of our iterated sparsification technique would also be of interest.

So far, the approach of constructing Cayley local spectral expanders over $\mathbb{F}_2^n$ by sparsifying Grassmann posets yields the sparsest known such complexes. In contrast to the success of this approach, we have limited understanding of its full power. To this end, we propose several questions: what codes can be used in place of the Hadamard codes so that the sparsification step preserves local spectral expansion? Could this approach be generalized to obtain local spectral expanders over other abelian groups? As mentioned above, we know that the approach we use can not give a complex of polynomial degree in n without introducing new ideas.

Coboundary expanders

Another fundamental question regards the isoperimetric inequalities we describe above. In this paper and many others, one uses the approach pioneered by Gromov [52] that applies isoperimetric inequalities to lower bounding coboundary expansion. A natural question is whether an isoperimetric inequality is also *necessary* to obtain coboundary expansion. An equivalence between coboundary expansion and isoperimetry will give us a simple alternative description for coboundary expansion. A counterexample to such a statement would motivate finding alternative approaches for showing coboundary expansion.

We call our family of 2-dimensional simplicial complexes that are both local spectral expanders and 1-dimensional coboundary expanders Johnson complexes. This construction can be generalized to yield families of k -dimensional Johnson complexes that are still local spectral expanders. However, it remains open whether they are also coboundary expanders in dimension > 1 .

1.5 Organization of this paper

In Section 2, we give preliminaries on simplicial complexes and local spectral expansion. We also define Grassmann posets—the vector space analogue for HDXs—which we use in our constructions. Other elementary backgrounds on graphs and expansion are given there as well.

In Section 3, we construct the subpolynomial degree Cayley complexes, proving Theorem 2. Most of this section discusses Grassmann posets instead of Cayley HDXs, but we describe the connection between the two given by [50] in this section too. We defer some of the more technical expansion upper bounds to Section D.

In Section 4 we construct the Johnson complexes which are both coboundary expanders and local spectral expanders. In this section we prove that they are local spectral expanders, leaving coboundary expansion to be the focus of the next two sections. We also give a detailed comparison of this construction to the one in [74], and discuss how to further sparsify our complexes by random subsampling.

In Section 5 we take a detour to formally define coboundary expansion and discuss its connection to isoperimetric inequalities. In this section we also give our version of the van-Kampen lemma, generalized to the setting of coboundary expansion. This may be of independent interest, as the van-Kampen lemma simplifies many proofs of coboundary expansion.

In Section 6 we prove that the Johnson complexes are coboundary expanders (Theorem 86), and local coboundary expanders (Theorem 89). We also prove that links in the construction of [50] are coboundary expanders.

In Section 7 we show that both the Johnson complex and the matrix complexes are special cases of a more general construction. In this section we show that the two complexes have a similar link structure, which is necessary for analyzing the local spectral expansion in both Cayley HDXs.

The appendices contain some of the more technical claims for ease of reading. In Section I we also give a lower bound on the degree of Cayley local spectral expanders, based on the degree of the link.

2 Preliminaries

We denote by $[k] = \{1, 2, \dots, k\}$.

2.1 Graphs and expansion

Let $G = (V, E, \mu)$ be a weighted graph where $\mu : E \rightarrow \mathbb{R}_{\geq 0}$ are the weights. In this paper we assume that μ is a probability distribution $\mu : E \rightarrow (0, 1]$ that gives every edge a positive measure. When μ is clear from context we do not specify it. We extend μ to a measure on vertices where $\mu(v) = \sum_{u \sim v} \mu(uv)$. When G is connected this is the stationary distribution of the graph. The adjacency operator of the graph A sends $f : V \rightarrow \mathbb{R}$ to $Af : V \rightarrow \mathbb{R}$ with $Af(v) = \mathbb{E}_{u \sim v} [f(u)] = \sum_{u \sim v} \frac{\mu(u)}{\mu(v)} f(u)$. We denote by $\langle f, g \rangle = \mathbb{E}_{v \sim \mu} [f(v)g(v)]$ the usual

inner product, and recall the classic fact that A is self adjoint with respect to this inner product. We denote by $\lambda(A)$ its second largest eigenvalue in absolute value. Sometimes we also write $\lambda(G)$ instead of $\lambda(A)$. It is well-known that $\lambda(G) \leq 1$.

► **Definition 8** (Expander graph). *Let $\lambda \geq 0$. A graph G is a λ -expander if $\lambda(G) \leq \lambda$.*

This definition is usually referred to as λ -two-sided spectral expansion. We note that the one-sided definition where one only gets an upper bound on the eigenvalues of G is also interesting, but is out of our paper's scope (see e.g. [56]).

When $G = (L, R, E)$ is a bipartite graph, we also have inner products on each side defined by $f, g : L \rightarrow \mathbb{R}$, $\langle f, g \rangle = \mathbb{E}_{v \in L} [f(v)g(v)]$ with respect to μ restricted to L (and similarly to real valued functions on R). The left-to-right bipartite graph operator A' that sends $f : L \rightarrow \mathbb{R}$ to $A'f : R \rightarrow \mathbb{R}$ with $A'f(u) = \mathbb{E}_{v \sim u} [f(v)]$ (resp. right-to-left bipartite operator). It is well known that $\|A'\|_{f \perp 1}$ is equal to the second largest eigenvalue of the non bipartite adjacency operator A defined above (*not* in absolute value). We also denote this quantity $\lambda(G) := \|A'\|_{f \perp 1}$ (it will be clear that whenever G is bipartite this is the expansion parameter in question).

► **Definition 9** (Bipartite expander graph). *Let $\lambda \geq 0$. A bipartite graph G is a λ -bipartite expander if $\lambda(G) \leq \lambda$.*

We use the following standard auxiliary statements.

▷ **Claim 10.** Let A, B be adjacency matrices of two graphs over the same vertex set and have the same stationary distribution. Let $\varepsilon = \max_v \|A_v - B_v\|_1 = \max_v \sum_{u \in V} |A(v, u) - B(v, u)|$. Then $\lambda(A) \leq \lambda(B) + \varepsilon$.

We will usually instantiate this claim with B being the complete graph with self loops, i.e. the edge distribution is just two independent choices of vertices. This B has $\lambda(B) = 0$ so we get this corollary.

► **Corollary 11.** *Let A be an adjacency matrix of a weighted graph $G = (V, E, \mu)$.*

1. $\lambda(A) \leq \max_{v \in V} \|A_v - \mu\|_1 = \sum_{u \in V} |A(v, u) - \mu(u)|$.
2. In particular, if A is uniform over a set $V' \subseteq V$ with $\mu(V') \geq 1 - p$ then $\lambda(A) \leq 2p$.

We also define a tensor product of graphs.

► **Definition 12.** *Let $G_1 = (V_1, E_1), G_2 = (V_2, E_2)$ be graphs. The tensor product $G_1 \otimes G_2$ is a graph with vertices $V_1 \times V_2$ and edges $(v_1, v_2) \sim (u_1, u_2)$ if $\{v_1, u_1\} \in E_1$ and $\{v_2, u_2\} \in E_2$. The measure on $G_1 \otimes G_2$ is the product of the original graphs' measures.*

We record the following well-known fact.

$$\lambda(G_1 \otimes G_2) = \max\{\lambda(G_1), \lambda(G_2)\}. \quad (2.1)$$

In fact, if $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ are the eigenvalues of G_1 and $\eta_1 \geq \eta_2 \geq \dots \geq \eta_m$ are the eigenvalues of G_2 , then the eigenvalues of $G_1 \otimes G_2$ are the multiset $\{\lambda_i \eta_j \mid i \in [n], j \in [m]\}$.

We also need the notion of a double cover.

► **Definition 13** (Double cover). *Let $G = (V, E)$ be a graph and $K_{1,1}$ be the graph that contains a single edge. The double cover of G is the graph $G \otimes K_{1,1}$. Explicitly, its vertex set is $V \otimes \{0, 1\}$ and we connect $\{(v, i), (u, j)\}$ if $\{v, u\} \in E$ and $i \neq j$.*

The following observation is well known.

► **Observation 14.** *A graph G is a λ -expander if and only if its double cover is a λ -bipartite expander.*

2.1.1 A local to global lemma

We will use the following “local-to-global” proposition due to [81, Theorem 1.1]. We prove it in Section A to be self contained. This lemma is an application of the Garland method [47], which is quite common in the high dimensional expander literature (see e.g. [87]). Let us set it up as follows.

Let $G = (V, E, \mu)$ be a graph where μ is the distribution over edges. A local-decomposition of G is a pair $\tau = (T, \nu_T)$ as follows. $T = \{G_t = (V_t, E_t, \mu_t)\}_{t \in T}$ is a set such that (V_t, E_t) are a subgraph of G and μ_t is a weight distribution over E_t . ν_T is a distribution over T . Assume that $\{v, u\} \sim \mu$ is sampled according to the following process:

1. Sample $t \sim \nu_T$.
2. Sample $\{v, u\} \sim \mu_t$.

The local-to-global graph $\mathcal{H}_\tau$ is the bipartite graph on V and T whose edge distribution is

1. Sample $t \sim \nu_T$.
2. Sample a vertex $v \sim \mu_t$ ⁷.
3. Output $\{v, t\}$.

Note that the probability of $v \in V$ under G and that under $\mathcal{H}_\tau$ (conditioned on the V side) are the same.

► **Lemma 15** ([81, Theorem 1.1]). *Let $G, \tau, \mathcal{H}_\tau$ be as above. Let $\lambda_2 = \lambda_2(\mathcal{H}_\tau)$ and $\gamma = \max\{\lambda(G_t) \mid t \in T\}$. Then $\lambda(G) \leq \gamma + \lambda_2^2(1 - \gamma)$.*

We will also use the one-sided version of [81, Theorem 1.1]. For a formal proof see [26, Lemma 4.14]. We formalize it in our language.

► **Lemma 16.** *Let $G = (L, R, E)$ be a bipartite graph. Let $\tau = (T, \nu_T)$ be a local decomposition. Let $\mathcal{H}_{\tau,L}$ be the restriction of $\mathcal{H}_\tau$ to the vertices of (L, T) (resp. $\mathcal{H}_{\tau,R}$). Then*

$$\lambda_2(G) \leq \max\{\lambda_2(G_t) \mid t \in T\} + \lambda_2(\mathcal{H}_{\tau,L})\lambda_2(\mathcal{H}_{\tau,R}).^8$$

2.2 Vector spaces

Unless stated otherwise, all vector spaces are over $\mathbb{F}_2$. For a vector space V and subspaces $u_1, u_2 \subseteq V$ we say their intersection is trivial if it is equal to $\{0\}$. We denote the sum $u_1 + u_2 = \{x_1 + x_2 \mid x_1 \in u_1, x_2 \in u_2\}$. If the sum is direct, that is the intersection of u_1 and u_2 is trivial, we write $u_1 \oplus u_2$. We shall use this notation to refer both to the statement that $u_1 \cap u_2 = \{0\}$ and to the object $u_1 + u_2$. We need the following claim.

► **Claim 17.** Let V be a vector space. Let $u_1, u_2, u_3 \subseteq V$ be three subspaces such that $(u_1 \oplus u_2) \cap u_3 = \{0\}$. Then $u_2 \cap (u_1 \oplus u_3) = \{0\}$.

Proof. Let $x_2 = x_1 + x_3 \in u_2 \cap (u_1 \oplus u_3)$ where $x_i \in u_i$. Thus $x_3 = x_2 - x_1 \in (u_1 \oplus u_2) \cap u_3$ which by assumption means that $x_3 = 0$ and $x_2 = x_1$. As $u_1 \cap u_2 = \{0\}$ it follows that $x_2 = x_1 = 0$ and the intersection is trivial. ◁

⁷ i.e. v is sampled with probability $\mu_t(v) = \frac{1}{2} \mathbb{P}_{e \sim \mu_t} [v \in e]$.

2.3 Grassmann posets

In this paper all vector spaces are with respect to $\mathbb{F}_2$ unless otherwise stated. Let V be a vector space. The Grassmann partially ordered set (poset) $\mathcal{G}(V)$ is the poset of subspaces of V ordered by containment. When V is clear from the context we sometimes just write $\mathcal{G}$.

► **Definition 18** (Grassman Subposet). *Let $n > 0$, V be an n -dimensional vector space, and $\ell \leq n$. An ℓ -dimensional Grassmann subposet Y is a subposet of $\mathcal{G}(\mathbb{F}_2^n)$ such that*

1. *every maximal subspace in Y has dimension ℓ ,*
2. *if $w \in Y$ and $w' \subseteq w$, then $w' \in Y$.*

We sometimes write $\dim(Y) = \ell$. The i -dimensional subspaces in Y are denoted $Y(i)$. The vector space V is sometimes called the ambient space of Y .⁹

A measured subposet is a pair (Y, μ) where Y is an ℓ -dimensional Grassmann subposet and μ is a probability distribution over ℓ -dimensional subspaces such that $\text{supp}(\mu) = Y(\ell)$. We extend μ to a distribution on *flags* of subspaces by sampling $v_\ell \sim \mu$ and then sampling a flag $(v_1 \subsetneq v_2 \subsetneq \dots \subsetneq v_\ell)$ uniformly at random. Sometimes we abuse notation and write $v_i \sim \mu$ where $v_i \in Y(i)$ is the marginal over said flags. We sometimes say that Y is measured without specifying μ for brevity. When we write $\mathcal{G}$ we mean that the distribution is uniform.

The containment graph is a bipartite graph with sides $(Y(i), Y(j))$ and whose edges $\{w, w'\}$ are such that $w \subseteq w'$. There is a natural distribution on the edges of this graph induced by the marginal of (v_i, v_j) of the flag above.

Let Y be an ℓ -dimensional Grassmann poset. For $i < \ell$ the Grassmann poset $Y^{\leq i} = \bigcup_{j=0}^i Y(j)$ is the poset that contains every subspace in Y of dimension $\leq i$.

A link of a subspace $w \in Y(i)$ is denoted Y_w . It consists of all subspaces in the poset that contain w , namely $Y_w = \{w' \supseteq w \mid w' \in Y\}$. If (Y, μ) is measured, the distribution over Y_w is the distribution μ conditioned on the sampled flag having $v_i = w$. We note that if V is the ambient space of Y then Y_w is (isomorphic to) a Grassmann subposet with ambient space V/w by the usual quotient map.

For any ℓ -dimensional subposet Y , any $i \leq \ell - 2$, $w \in Y(i)$, and $i < j \leq \ell$, we denote $Y_w(j) = \{w' \supseteq w \mid w' \in Y(j)\}$. The underlying graph of the link is the following graph (which we also denote by Y_w when there is no confusion). The vertices are

$$U_w = \{v \in V \mid \text{span}(v) + w \in Y(i+1)\}$$

and the edges are

$$E_w = \{\{v, v'\} \mid \text{span}(v, v') + w \in Y(i+2)\}.$$

We note that sometimes the link graph is defined by taking the vertices to be $Y_w(i+1)$ and the edges to be all $\{w', w''\}$ such that $w' + w'' \in Y(i+2)$. These two graphs are closely related:

▷ **Claim 19.** Let Y be a Grassmann poset. Let $w \in Y(i)$ be a subspace and let $Y_w = (U_w, E_w)$ be the underlying graph of its link as defined above. Let $G = (V, E)$ be the graph whose vertices are $V = Y_w(i+1)$ and whose edges are $\{w', w''\}$ such that $w' + w'' \in Y(i+2)$. Then $\lambda(Y_w) = \lambda(G)$.

⁹ formally there could be many such spaces, but in this paper we usually assume that V is spanned by the subspaces in Y .

Proof. Let K_{2^i} be the complete graph over 2^i vertices including self loops. We prove that $G \cong Y_w \otimes K_{2^i}$. As $\lambda(K_{2^i}) = 0$ it will follow by (2.1) that $\lambda(G) = \max\{\lambda(Y_w), 0\} = \lambda(Y_w)$. For every $w' \in Y_w(i+1)$ we order the vectors in $w' \setminus w$ according to some arbitrary ordering $w' \setminus w = \{v_1, v_2, \dots, v_{2^i}\}$. Our isomorphism from G to $Y_w \otimes K_{2^i}$ sends $v_j \mapsto (w', j)$ where $w' = w + \text{span}(v_j)$ and the j in the subscript is according to the ordering of $w' \setminus w$. It is direct to check that this is a bijection.

Moreover, $v \sim v'$ in G if and only if $w + \text{span}(v, v') \in Y(i+2)$. This span is equal to $w + \text{span}(v, v') = (w + \text{span}(v)) + (w + \text{span}(v'))$. Therefore $v \sim v'$ if and only if the left coordinates of the images $(w + \text{span}(v), j_1)$ and $(w + \text{span}(v'), j_2)$ are connected in Y_w . As $j_1 \sim j_2$ is an edge in the complete graph with self loops, it follows that $v \sim v'$ is in G if and only if the corresponding images are connected in $Y_w \otimes K_{2^i}$. The claim follows. $\triangleleft$

Thus we henceforth use our definition of a link.

We say that such a subposet is a λ -expander if for every $i \leq \ell - 2$ and $w \in Y(i)$ the graph Y_w is a λ -two-sided spectral expander. We also state the following claim, whose proof is standard.

▷ **Claim 20.** The poset $Y = \mathcal{G}(\mathbb{F}_2^n)^{\leq d}$ is a $\frac{1}{2^{n-d}-2}$ -expander.

Sketch. Let $w \in Y(i)$. By Claim 19, to analyze the expansion of the link of w , it suffices to consider the graph whose vertices are the subspaces $w' \in Y_w(i+1)$ and whose edges are $w' \sim w''$ if $w' + w'' \in Y_w(i+2)$. In this case, it is easy to see that this graph is just the complete graph (without self loops) with $2^{n-i} - 1$ vertices. The claim follows by the expansion of the complete complex. $\triangleleft$

We also need the following claim.

▷ **Claim 21 ([30]).** Let Y be a Grassmann poset that is λ -expander. Then for every $i \leq j$ and $w \in Y$, the containment graph between $Y_w(i)$ and $Y_w(j)$ is a $(0.61 + j\lambda)^{\frac{1}{2}(j-i)}$ -bipartite expander.

This claim was proven [30] without the explicit constant 0.61. Hence we reprove it in Section H with the constant.

2.4 The matrix domination poset

In this section we give preliminaries on the matrix domination posets. These posets are essential for proving that the HDX construction given in Section 3 sufficiently expands. They were a crucial component in the construction of abelian HDXs in [50]. We believe that the matrix domination posets are also interesting on their own right, being another important family of posets to be studied as those in [30, 69].

Let $\mathcal{M}$ be $n \times n$ matrices over $\mathbb{F}_2$ for some fixed n . Let $\mathcal{M}(i) = \{A \in \mathcal{M} \mid \text{rank}(A) = i\}$. The partial order on this set is the domination order.

► **Definition 22** (Matrix domination partial order). *Let $A, B \in \mathcal{M}$. We say that A dominates B if $\text{rank}(A) = \text{rank}(B) + \text{rank}(A - B)$. In this case we write $B \leq A$.*

Another important notion is the direct sum.

► **Definition 23** (Matrix direct sum). *We say that two matrices A and B have a direct sum if $\text{rank}(A) + \text{rank}(B) = \text{rank}(A + B)$. In this case we write $A \oplus B$.*

We note that this definition is equivalent to saying that $A, B \leq A + B$. We also clarify that when we write $A \oplus B$, this refers both to the object $A + B$, and is a statement that $\text{rank}(A) + \text{rank}(B) = \text{rank}(A + B)$. This is similar to the use of $\oplus$ when describing a sum of subspaces $U_1, U_2 \subseteq \mathbb{F}_2^n$; the term $U_1 \oplus U_2$ refers both to the sum of subspaces, and is a statement that $U_1 \cap U_2 = \{0\}$. We will see below that $A \oplus B$ is equivalent to $\text{row}(A) \cap \text{row}(B) = \{0\}, \text{col}(A) \cap \text{col}(B) = \{0\}$, so we view this notation as a natural analogue of what happens in subspaces.

In Section 3 we define and analyze expansion of graphs coming from Hasse diagrams of dominance orders. These graphs appear in the analysis of the construction in Section 3. Before doing so, let us state some preliminary properties of this partial order, all of which are proven in Section C. More specifically, we will show that this is a valid partial order (as already observed in [50]), discuss the properties of pairs of matrices that have a direct sum, and study structures of *intervals* in this poset (see definition below).

▷ **Claim 24.** The pair $(\mathcal{M}, \leq)$ is a partially ordered set.

We move on to characterize when $A \oplus B$, that is, when the sum of A and B is direct. This will be important in the subsequent analyses.

▷ **Claim 25.** For any matrices A, B the following are equivalent:

1. $\text{rank}(A + B) = \text{rank}(A) + \text{rank}(B)$, i.e. $A \oplus B$.
2. The spaces $\text{row}(A) \cap \text{row}(B)$ and $\text{col}(A) \cap \text{col}(B)$ are trivial.
3. The spaces $\text{col}(A + B) = \text{col}(A) \oplus \text{col}(B)$ and $\text{row}(A + B) = \text{row}(A) \oplus \text{row}(B)$.
4. There exists a decomposition $A = \sum_{j=1}^{\text{rank}(A)} e_j \otimes f_j$ and $B = \sum_{j=1}^{\text{rank}(B)} h_j \otimes k_j$ such that the disjoint unions $\{e_j\} \cup \{h_j\}$ and $\{f_j\} \cup \{k_j\}$ are independent sets. Indeed, the two sets span $\text{row}(A + B)$ and $\text{col}(A + B)$ respectively.

We will also use this elementary (but important) observation throughout.

▷ **Claim 26 (Direct sum is associative).** Let $A, B, C \in \mathcal{M}$. If A and B have a direct sum, and C has a direct sum with $A \oplus B$, then A has a direct sum with $C \oplus B$. Stated differently, if $A \oplus B$ and $C \oplus (A \oplus B)$, then $A \oplus (C \oplus B)$. The same holds for any other permutation of A, B and C .

Henceforth, we just write $A_1 \oplus A_2 \oplus A_3$ to indicate that $A_i \oplus (A_j \oplus A_k)$ for any three distinct i, j, k .

Finally, we will analyze subposets of $\mathcal{M}$ that correspond to local components which are analogous to the links in simplicial complexes and Grassmann posets.

► **Definition 27 (Intervals).** Let $M_1 \leq M_2$ be two matrices. The interval between M_1 and M_2 is the subposet of $\mathcal{M}$ induced by the set

$$\mathcal{M}_{M_1}^{M_2} = \{A \mid M_1 \leq A \leq M_2\}.$$

Similar to the above, we also denote by $\mathcal{M}_{M_1}^{M_2}(i)$ the matrices in $\mathcal{M}_{M_1}^{M_2}$ of rank i . When $M_1 = 0$ we just write $\mathcal{M}^{M_2}$.

The main observation we need on these intervals is this.

► **Proposition 28.** Let $M_1 \leq M_2$ be of ranks $m_1 \leq m_2$. The subposets $\mathcal{M}_{M_1}^{M_2}$ and $\mathcal{M}^{M_2 - M_1}$ are isomorphic by $A \mapsto A' = A - M_1$. This isomorphism has the property that

$$\text{rank}(A') = \text{rank}(A) - m_1.$$

Another subposet we will be interested in is the subposet of matrices that are “disjoint” from some $A \in \mathcal{M}$. That is,

► **Definition 29.** Let $A \in \mathcal{M}$. The subposet $\mathcal{M}_{\cap A}$ is the subposet of $\mathcal{M}$ consisting of all $\{B \in \mathcal{M} \mid B \oplus A\}$.

We note that by Claim 25 $\mathcal{M}_{\cap A} = \{B \in \mathcal{M} \mid \text{row}(B) \cap \text{row}(A), \text{col}(B) \cap \text{col}(A) = \{0\}\}$.

▷ **Claim 30.** Let $A \in \mathcal{M}(i)$. Then $\phi : \mathcal{M}_A \rightarrow \mathcal{M}_{\cap A}$, $\phi(B) = B - A$ is an order preserving isomorphism.

2.5 Simplicial complexes and high dimensional expanders

A pure d -dimensional simplicial complex X is a hypergraph that consists of an arbitrary collection of sets of size $(d+1)$ together with all their subsets. The i -faces are sets of size $i+1$ in X , denoted by $X(i)$. We identify the vertices of X with its singletons $X(0)$ and for good measure always think of $X(-1) = \{\emptyset\}$. The k -skeleton of X is $X^{\leq k} = \bigcup_{i=-1}^k X(i)$. In particular the 1-skeleton of X is a graph. We denote by $\text{diam}(X)$ the *diameter* of the graph underlying X .

2.5.1 Probability over simplicial complexes

Let X be a simplicial complex and let $\mathbb{P}_d : X(d) \rightarrow (0, 1]$ be a density function on $X(d)$ (that is, $\sum_{s \in X(d)} \mathbb{P}_d(s) = 1$). This density function induces densities on lower level faces $\mathbb{P}_k : X(k) \rightarrow (0, 1]$ by $\mathbb{P}_k(t) = \frac{1}{\binom{d+1}{k+1}} \sum_{s \in X(d), s \supseteq t} \mathbb{P}_d(s)$. We omit the level of the faces when it is clear from the context, and just write $\mathbb{P}[T]$ or $\mathbb{P}_{t \in X(k)}[T]$ for a set $T \subseteq X(k)$.

2.5.2 Links and local spectral expansion

Let X be a d -dimensional simplicial complex and let $s \in X$ be a face. The link of s is the $d' = d - |s|$ -dimensional complex

$$X_s = \{t \setminus s \mid t \in X, t \supseteq s\}.$$

For a simplicial complex X with a measure $\mathbb{P}_d : X(d) \rightarrow (0, 1]$, the induced measure on $\mathbb{P}_{d', X_s} : X_s(d - |s|) \rightarrow (0, 1]$ is

$$\mathbb{P}_{d', X_s}(t \setminus s) = \frac{\mathbb{P}_d(t)}{\sum_{t' \supseteq s} \mathbb{P}_d(t')}.$$

► **Definition 31** (Local spectral expander). Let X be a d -dimensional simplicial complex and let $\lambda \geq 0$. We say that X is a λ -local spectral expander if for every $s \in X$, the link of s is connected, and for every $s \in X(d-2)$ it holds that $\lambda(X_s) \leq \lambda$.

We stress that this definition includes connectivity of the underlying graph of X , because of the face $s = \emptyset$. In addition, in some of the literature the definition of local spectral expanders includes a spectral bound $\lambda(X_s) \leq \lambda$ for all $i \leq d-2$ and $s \in X(i)$. Even though the definition we use only requires a spectral bound for links of faces $s \in X(d-2)$, it is well known that when λ is sufficiently small, this actually implies bounds on $\lambda(X_s)$ for all $i \leq d-2$ and $s \in X(i)$ (see [87] for the theorem and more discussion).

Cayley Graphs and Cayley Local Spectral Expanders

► **Definition 32** (Cayley graph over $\mathbb{F}_2^n$). Let $S \subseteq \mathbb{F}_2^n$. A Cayley graph over $\mathbb{F}_2^n$ with generating set S is a graph whose vertices are $\mathbb{F}_2^n$ and whose edges are $E = \{\{x, x+s\} \mid x \in \mathbb{F}_2^n, s \in S\}$. We denote such a graph by $\text{Cay}(\mathbb{F}_2^n, S)$.

Note that we can label every edge in E by its corresponding generator. A Cayley complex is a simplicial complex X such that $X^{\leq 1}$ is a Cayley graph (as a set) and such that all edges with the same label have the same weight. We say that X is a λ -two-sided (one-sided) Cayley local spectral expander if it is a Cayley complex and a λ -two-sided (one-sided) local spectral expander.

3 Subpolynomial degree Cayley local spectral expanders

In this section we construct Cayley local spectral expander with a subpolynomial degree, improving a construction by [50].

► **Theorem 33.** For every $\lambda, \varepsilon > 0$ and integer $d' \geq 2$ there exists an infinite family $\{X_k\}_{k=1}^\infty$ of d' -dimensional λ -local spectral expanding Cayley complexes with $X(0) = \mathbb{F}_2^{n_k}$. The degree is upper bounded by $2^{(n_k)^\varepsilon}$.

Golowich proved this theorem for $\varepsilon = \frac{1}{2}$.

As in Golowich's construction, we proceed as follows. We first construct an expanding family of Grassmann posets $\{Y_k\}_{k=1}^\infty$ over $\mathbb{F}_2$ (see Section 2.3), and then we use a proposition by [50] (Proposition 37 below) to construct said Cayley local spectral expanders. The main novelty in this part of the paper is showing that this idea can be iterated, leading to a successive reduction of the degree.

The following notation will be useful later.

► **Definition 34.** Let d, n, D be integers and $\lambda > 0$. We say a Grassmann poset is a $[d, n, D, \lambda]$ -poset if Y is a d -dimensional λ -local spectral expander whose ambient space is $\mathbb{F}_2^n$, with $|Y(1)| \leq 2^D$.

[50] introduces a general approach that turns a Grassmann poset Y into a simplicial complex $\beta(Y)$ called the basification of Y . The formal definition of this complex is as follows.

► **Definition 35.** Given a d -dimensional Grassmann poset Y . Its basification $\beta(Y)$ is a $(d-1)$ -dimensional simplicial complex such that for every $0 \leq i \leq d-1$

$$\beta(Y)(i) = \{\{v_0, \dots, v_i\} \mid \text{span}(v_0, \dots, v_i) \in Y(i+1)\}.$$

In [50] it is observed that the local spectral expansion of Y is identical to that of $\beta(Y)$. Golowich constructs a Cayley complex whose vertex links are (copies of) the basification. The construction is given below.

► **Definition 36.** Given the basification $\beta(Y)$ of a d -dimensional Grassmann complex Y over the ambient space $\mathbb{F}_2^n$, define the d -dimensional abelian complex $X = \text{Cay}(\mathbb{F}_2^n, \beta(Y))$ such that $X(0) = \mathbb{F}_2^n$ and for $0 < i \leq d$

$$X(i) = \{\{x, x+v_1, x+v_2, \dots, x+v_i\} \mid \{v_1, v_2, \dots, v_i\} \in \beta(Y)(i-1)\}.$$

We observe that the underlying graph is $\text{Cay}(\mathbb{F}_2^n, Y(1))$ and therefore its degree is 2^D .

► **Proposition 37** ([50]). *Let Y be a d -dimensional Grassmann poset over $\mathbb{F}_2^n$ that is λ -expanding. Let X be the abelian complex given by its basification. Then X is a λ -local spectral expander. In addition, the underlying graph of X is a $\frac{\lambda}{1-\lambda}$ -expander.*

Thus our new goal is to prove the following theorem about expanding posets.

► **Theorem 38.** *For every $\varepsilon, \lambda' > 0$, and d' there exists infinite family of integers n'_k and of $[d', n'_k, n_k^{\varepsilon}, \lambda']$ -posets Y_k .*

This is sufficient to prove Theorem 33:

Proof of Theorem 33, assuming Theorem 38. Let Y_k be an infinite family of $[d, n_k, n_k^{\varepsilon}, \lambda']$ in Theorem 38 for λ' such that $\lambda' = \frac{\lambda'}{1-\lambda'}$. Let X_k be their basifications. By Proposition 37. The complexes X_k have ambient subspace $\mathbb{F}_2^{n_k}$, degree $|Y(1)| \leq 2^{(n_k)^{\varepsilon}}$, and expansion λ' . ◀

3.1 Expanding Grassmann posets from matrices

Let Y be a $[d, n, D, \lambda]$ -poset. We now describe how to construct from it a $[d', n', D', \lambda']$ -poset Y' where $d' = \Omega(\log d)$, $n' = n^2$, $D' = d \cdot D$ and $\lambda' = 2^{-\Omega(d')}$ when $\lambda \leq \frac{1}{4d}$. In other words, we describe a transformation

$$[d, n, D, \lambda]\text{-poset } Y \longmapsto [d', n', D', \lambda']\text{-poset } Y'.$$

After this, we will iterate this construction, this time taking Y' to be the input, producing an even sparser poset Y'' etc., as we illustrate further below.

Without loss of generality we identify the ambient space of Y with $\mathbb{F}_2^n$ and set the ambient space of Y' to be $n \times n$ matrices ($\cong \mathbb{F}_2^{n^2}$). The 1-dimensional subspaces of Y' are

$$Y'(1) = \left\{ M \mid \text{rank}(M) = \frac{d}{2} \text{ and } \text{row}(M), \text{col}(M) \in Y\left(\frac{d}{2}\right) \right\}. \quad (3.1)$$

In words, these are all matrices whose rank is $d/2$ so that their row and column spaces are in Y . As we can see, $n' = n^2$ and $|Y'(1)| \leq |Y(1)|^d$ and therefore $D' = \log |Y'(1)| \leq dD$. We postpone the construction higher dimensional faces (which determine the dimension d') and the intuition behind the expansion of Y' for later. Let us first show how an iteration of this construction will construct sparser Grassmann posets.

This single step can be iterated any constant number of times, starting from a high enough initial dimension d_0 . Let us denote by $\mathcal{SP}(Y)$ the operator that takes as input one Grassmann poset Y and outputs the sparser one Y' as above. Fix m and fix a target dimension d_m . Let $d_0 \gg d_m$ and let n_0 be sufficiently large. We set $Y_0, Y_1, \dots, Y_m$ to be such that $Y_0 = \mathcal{G}(\mathbb{F}_2^{n_0})^{\leq d_0}$ (which is a $[d_0, n_0, D_0, \lambda_0]$ -poset for $D_0 = n_0$ and λ_0 that goes to 0 with n_0) and $Y_{i+1} = \mathcal{SP}(Y_i)$. The final poset Y_m is a

$$[d_m, n_m, D_m, \lambda_m]\text{-poset}.$$

Here $d_m = \Omega_m(\underbrace{\log \circ \log \circ \dots \circ \log}_{m \text{ times}} d_0)$, $n_m = n_0^{2^m}$, $D_m = \tilde{O}(d) \cdot n$ and $\lambda_m = 2^{-\Omega(d_m)}$ when $\lambda_0 = 2^{-\Omega(d_0)}$.

The above construction works for any large enough n_0 . This gives us an infinite family of Grassmann complexes whose ambient dimension is $n_m = n_0^{2^m}$ and whose degree is $D_m = n_m^{\varepsilon}$ (where $m = \log \frac{1}{\varepsilon}$). This infinite family proves Theorem 38.

3.1.1 The construction of the Grassmann poset $\mathcal{SP}(Y)$

We are ready to give the formal definition of the poset. We construct the top level subspaces, and then take their downward closure to be our Grassmann poset. Every top level face is going to be an image of a “non-standard” Hadamard encoding. The “standard” Hadamard could be described as follows. Let $\mathbb{F}_2^i$ be a vector space, and let $\mathbb{F}_2^{2^i}$ be indexed by the vectors in $\mathbb{F}_2^i$. That is $\mathbb{F}_2^{2^i} \cong \text{span}(\{e_v\}_{v \in \mathbb{F}_2^i})$ where e_v is the standard basis vector that is 1 in the v -th coordinate and 0 elsewhere. With this notation the “standard” Hadamard encoding maps

$$x \mapsto \sum_{v \in \mathbb{F}_2^i} \langle x, v \rangle e_v$$

where $\langle x, v \rangle = \sum_{i=1}^n x_i v_i \pmod{2}$ is the standard bilinear form. A “non-standard” Hadamard encoding in this context is the same as above, only replacing the standard vectors e_v with another basis for a subspace, represented by $2^i \times n$ matrices. It is defined by a set $\{s(v)\}_{v \in \mathbb{F}_2^i}$ satisfying some properties described formally below. The encoding is the linear map $\hat{s} : \mathbb{F}_2^i \rightarrow \mathbb{F}_2^{n \times n}$,

$$\hat{s}(x) = \sum_{v \in \mathbb{F}_2^i} \langle x, v \rangle s(v)$$

similar to the formula above. We continue with a formal definition.

We start with Y , a $[d, n, D, \lambda]$ -poset (and assume without loss of generality that d is a power of 2) and let $d' = \log d^{10}$.

► **Definition 39** (Admissible function). *Consider a $[d, n, D, \lambda]$ -poset Y with $d' = \log d$. Let $i > 0$, a (not necessarily linear) function $s : \mathbb{F}_2^i \setminus \{0\} \rightarrow \mathbb{F}_2^{n \times n}$ is admissible to Y if:*

1. *For every $v \in \mathbb{F}_2^i \setminus \{0\}$, $\text{rank}(s(v)) = 2^{d'-i}$.*
2. *The sum matrix $M_s = \bigoplus_{v \in \mathbb{F}_2^i \setminus \{0\}} s(v)$ is a direct sum of the $s(v)$'s, that is,*

$$\text{rank}(M_s) = \sum_{v \in \mathbb{F}_2^i \setminus \{0\}} \text{rank}(s(v)) = 2^{d'} - 2^{d'-i}.$$

3. *The spaces $\text{col}(M_s), \text{row}(M_s) \in Y(2^{d'} - 2^{d'-i})$.*

We remark that for the construction we only need to consider admissible functions for $i = d'$, that is, functions $s : \mathbb{F}_2^{d'} \setminus \{0\} \rightarrow \mathbb{F}_2^{n \times n}$, where $\text{rank}(s(v)) = 1$, $\text{rank}(M_s) = 2^{d'} - 1$ and $\text{col}(M_s), \text{row}(M_s) \in Y(2^{d'} - 1)$. The other admissible functions will be used later in the analysis (see Section 7).

We denote by $\mathcal{S}_i$ the set of admissible functions $s : \mathbb{F}_2^i \setminus \{0\} \rightarrow \mathbb{F}_2^{n \times n}$. The *Hadamard encoding* of a function $s : \mathbb{F}_2^i \setminus \{0\} \rightarrow \mathbb{F}_2^{n \times n}$ is the linear function $\hat{s} : \mathbb{F}_2^i \rightarrow \mathbb{F}_2^{n \times n}$ given by

$$\hat{s}(x) = \sum_{v \in \mathbb{F}_2^i \setminus \{0\}} \langle x, v \rangle s(v),$$

Observe that $\hat{s}$ is an isomorphism so the image of $\hat{s}$ is an i -dimensional subspace. This of course means that every subspace $u \subseteq \mathbb{F}_2^i$ is mapped isomorphically to a subspace of the same dimension $\text{Im}(\hat{s}|_u)$. Thus, the Grassmann poset $\mathcal{G}_s := \{\text{Im}(\hat{s}|_u) \mid u \subseteq \mathbb{F}_2^i, u \text{ is a subspace}\}$ is isomorphic to the complete Grassmann $\mathcal{G}(\mathbb{F}_2^i)$.

¹⁰ Although we define this poset to the highest dimension possible. Our construction will eventually take a lower dimensional skeleton of it.

► **Definition 40** (Matrix poset). *The poset $Y' = \mathcal{SP}(Y)$ is a d' -dimensional poset $Y' = \bigcup_{s \in \mathcal{S}_{d'}} \mathcal{G}_s$. We denote by $\mathcal{SP}(Y)$ the operator that takes Y and produces Y' as above.*

In other words, Y' is clearly a Grassmann poset and it is a union of $\mathcal{G}_s$ for all admissible mappings $s \in \mathcal{S}_{d'}$.

The admissible functions on level i give us a natural description for $Y'(i)$:

▷ **Claim 41.** Let Y' be as in Definition 40. Then

$$Y'(i) = \{Im(\hat{s}) \mid s \in \mathcal{S}_i\}.$$

We prove this claim in Section B.

This poset also comes with a natural distribution over d' -subspaces is the follows.

1. Sample two $(d-1)$ -dimensional subspaces $w_1, w_2 \in Y(d-1)$ independently according to the distribution of Y .
2. Sample uniformly an admissible function $s \in \mathcal{S}_{d'}$ such that the matrix $M_s = \bigoplus_{v \in \mathbb{F}_2^{d'} \setminus \{0\}} s(v)$ has $\text{row}(M_s) = w_1$ and $\text{col}(M_s) = w_2$.
3. Output $Im(\hat{s})$.

We remark that we can generalize this construction by changing the definition of “admissible” to other settings. One such generalization that replaces the rank function with the hamming weight results in the Johnson poset which we describe in Section 4. We explore this more generally in Section 7.

We finally reach the main technical theorem of this section, Theorem 42. This theorem proves that if Y is an expanding poset, then so is a skeleton of $\mathcal{SP}(Y)$. Using this theorem inductively is the key to proving Theorem 38.

► **Theorem 42.** *There exists a constant $K > 0$ such that the following holds. Let $\lambda' \in (0, 1)$, $D > 0$ and $d' > 0$ be an integer. Let $d = \max\{2^{2d'}, (K \log \frac{1}{\lambda'})^2\}$ and let $\lambda \leq \frac{1}{4d}$. If Y is a $[d, n, D, \lambda]$ -poset then the d' -skeleton of $Y' = \mathcal{SP}(Y)$, namely $Y'^{\leq d'}$, is a $[d', n', D', \lambda']$ -poset where $n' = n^2$ and $D' = d \cdot D$.*

Proof of Theorem 38, assuming Theorem 42. Fix ε and without loss of generality $\varepsilon = \frac{1}{2^m}$ for some integer $m \geq 0$. Changing variables, it is enough to construct a family $\{Y_k\}$ of $[d', n_k^{2^m}, O(n_k), \lambda']$ -posets for an infinite sequence of integers n_k , such that the constant in the big O only depends on d', λ' and m but not on n_k . Then by setting $n'_k = n_k^{2^m}$ gives us the result ¹¹.

We prove this by induction on m for all d' and λ' simultaneously. For $m = 0$ we take the complete Grassmanns, namely $n_k = k$ and the family $Y = Y_k = \mathcal{G}(\mathbb{F}_2^k)^{\leq d'}$ that are $[d', k, k, \lambda']$ -posets for sufficiently large k (the first three parameters are clear. By Claim 20 the links expansion of the complete Grassmann is $\leq \frac{1}{2^{(k-d')}-1}$ which goes to 0 with k , hence for every $\lambda' > 0$ and large enough k they are λ' -expanders).

Let us assume the induction hypothesis holds for m and prove it for $m+1$. Fix $\lambda' > 0$ and d' . By Theorem 42, there exists some d and λ such that if we can find a family of $[d, n_k^{2^m}, O(n_k), \lambda]$ -posets $\{Y_k\}_{k=1}^\infty$, then the posets $Y'_k := \mathcal{SP}(Y_k)^{\leq d'}$ are $[d', n_k^{2^{m+1}}, O(n_k \cdot d), \lambda']$ -posets. By the induction hypothesis (applied for d and λ) such a family $\{Y_k\}_{k=1}^\infty$ indeed exists. The theorem is proven. ◀

¹¹Technically this only gives a family $[d', n'_k, O(n'_k), \lambda']$ -posets. But $\varepsilon > 0$ can be arbitrarily small. Thus for any $\varepsilon > 0$, constructing $[d', n', C \cdot n'^{\varepsilon/2}, \lambda']$ -posets is also a construction of $[d', n'_k, n'_k, \lambda']$ -posets, since eventually $n'_k > C \cdot n'^{\varepsilon/2}$ for any constant C independent of n'_k .

3.2 A Proof Roadmap for a single iteration of $\mathcal{SP}$ - Theorem 42

The rest of this section is devoted to proving Theorem 42, and mainly to bounding the expansion of links in Y' . Recall the notation $\mathcal{M}$ to be the poset of matrices with the domination relation.

We fix the parameters λ' and d' . Without loss of generality we take d' to be large enough so that $d = 2^{2d'}$ (this requires that $d' \geq \log K + \log \log \frac{1}{\lambda'}$ for the constant $K > 0$ in Theorem 42).

Recall the partial order defined in Section 2.4 on matrices where we write $A \leq M$ for two matrices $A, M \in \mathbb{F}_2^{n \times n}$ if $\text{rank}(M) = \text{rank}(A) + \text{rank}(M - A)$.

Fix Y , and let $Y' = \mathcal{SP}(Y)$. Let $W \in Y'(\ell)$ be a subspace (where $\ell \leq d'$). In the next two subsections we describe the two steps of the proof: decomposition and upper bounding expansion.

3.2.1 Step I: decomposition

We will show that a link of a subspace $W \in Y'(\ell)$ decomposes to a tensor product of simpler components.

For this we must define two types of graphs, relying on the domination relation of matrices and the direct sum.

► **Definition 43** (Subposet graph). *Let $m > 0$ and let $U \in \mathcal{M}(4m)$. The graph $\mathcal{T}^U = (V, E)$ has vertices $V = \mathcal{M}^U(2m)$ and edges $\{A, B\}$ such that there exists $C \in \mathcal{M}(m)$ such that $C \oplus (A - C) \oplus (B - C) \in \mathcal{M}^U(3m)$.*

For any two $U_1, U_2 \in \mathcal{M}(4m)$ of rank $4m$, the graphs $\mathcal{T}^{U_1} \cong \mathcal{T}^{U_2}$ so we denote this graph by $\mathcal{T}^m$ when we do not care about the specific matrix.

► **Definition 44** (Disjoint graph). *Let $m, i > 0$ and let $D \in \mathcal{M}(i)$. The graph $\mathcal{H}_{D,m} = (V, E)$ has vertices $V = \mathcal{M}_{\cap D}(2m)$ and edges $\{A, B\}$ such that there exists $C \in \mathcal{M}(m)$ such that we have $C \oplus (A - C) \oplus (B - C) \in \mathcal{M}_{\cap D}(3m)$.*

For any n and two $D_1, D_2 \in \mathcal{M}(i)$, the graphs $\mathcal{H}_{D_1,m} \cong \mathcal{H}_{D_2,m}$ so we denote this graph by $\mathcal{H}_{i,m}$ when we do not care about the specific matrix.

The first component in the decomposition in the link of W is the subposet graph. The second component is the sparsified link graph. This graph depends on Y , unlike the previous component. In particular, the matrices in this component will have row and column spaces in Y . For the rest of this section we fix $m = \frac{d}{2^{\ell+2}}$. Let Z_W be any matrix whose row span and column span are $\text{row}(Z_W) = \sum_{M \in W} \text{row}(M)$ and $\text{col}(Z_W) = \sum_{M \in W} \text{col}(M)$. Note that both $\text{row}(Z_W)$ and $\text{col}(Z_W)$ have dimension $(1 - \frac{1}{2^\ell})d$ (that is, $d - 4m$). The following definition will only depend on these row and column spaces, but it will be easier in the analysis to give the definition using this matrix Z_W .

► **Definition 45** (Sparsified link graph). *The graph $\mathcal{L}_{Y,W}$ is the graph with vertex set*

$$V = \{A \in \mathcal{M}_{\cap Z_W}(2m) \mid \text{row}(A) \oplus \text{row}(Z_W), \text{col}(A) \oplus \text{col}(Z_W) \in Y(d - 2m)\}.$$

The edges are all $\{A_1, A_2\}$ such that there exists a matrix $U \in \mathcal{M}_{\cap Z_W}(4m)$ such that $\text{row}(U) \oplus \text{row}(Z_W), \text{col}(U) \oplus \text{col}(Z_W) \in Y(d)$, and $\{A_1, A_2\}$ are an edge in $\mathcal{T}^U$.

Formally, the distribution over edges in the sparsified link graph $\mathcal{L}_{Y,W}$ is the following.

1. Sample $w_1 \in Y_{\text{row}(Z_W)}(d)$ and $w_2 \in Y_{\text{col}(Z_W)}(d)$ according to their respective distributions.
2. Uniformly sample a matrix U such that $\text{row}(Z_W) \oplus \text{row}(U) = w_1$ and $\text{col}(Z_W) \oplus \text{col}(U) = w_2$.
3. Sample an edge in $\mathcal{T}^U$.

We can now state our decomposition lemma:

► **Lemma 46.** $Y'_W \cong \underbrace{(\mathcal{T}^m \otimes \mathcal{T}^m \otimes \cdots \otimes \mathcal{T}^m)}_{2^\ell - 1 \text{ times}} \otimes \mathcal{L}_{Y,W}.$

This lemma is proven in Section 3.3.

3.2.2 Step II: upper bounding expansion

It remains to show that $\mathcal{T}^m$ and $\mathcal{L}_{Y,W}$ are expander graphs. We prove the following two lemmas.

► **Lemma 47.** *There exists an absolute constant $\gamma \in (0, 1)$ such that $\lambda(\mathcal{T}^m) \leq \gamma^m$ for all $m \leq d$.*

In particular, there is an absolute constant $K_1 > 0$ such that the following holds: Let $\lambda' > 0$. If $d' \geq K_1 + \log \log \frac{1}{\lambda'}$ and $d \geq 2^{2d'}$, then

$$\lambda(\mathcal{T}^m) \leq \lambda'.$$

► **Lemma 48.** *There exists a constant $\gamma \in (0, 1)$ such that the following holds. Let Y be a $[d, n, D, \lambda]$ -poset for $\lambda < \frac{1}{4d}$, and let $W \in Y'(\ell)$ where $Y' = \mathcal{SP}(Y)^{\leq d'}$ and $\ell \leq d'$. Then $\lambda(\mathcal{L}_{Y,W}) \leq \gamma^m$.*

In particular, there is an absolute constant $K_2 > 0$ such that the following holds: Let $\lambda' > 0$. If $d' \geq K_2 + \log \log \frac{1}{\lambda'}$, $d \geq 2^{2d'}$ and $\lambda \leq \frac{1}{4d}$, then

$$\lambda(\mathcal{L}_{Y,W}) \leq \lambda'.$$

The constant K in Theorem 42 is chosen to be $2^{\max(K_1, K_2)}$. We now give an high level explanation of the proofs of Lemma 47 and Lemma 48, and afterwards we use them to prove Theorem 42.

The graph $\mathcal{T}^m$ does not depend on Y and the proof of Lemma 47 relies only on the properties of the matrix domination poset. Therefore it is deferred to Section D. In a bird's-eye view, we use Lemma 16, the decomposition lemma, to decompose the graph into smaller subgraphs, and show that every subgraph is close to a uniform subgraph.

We prove Lemma 48 in Section 3.4. Its proof also revolves around Lemma 16, only this time the subgraphs in the decomposition come from (pairs of) subspaces in Y . Observe that in the first step in sampling an edge in Definition 45, one first samples a pair $(w_1, w_2) \in Y_{\text{row}(Z_W)}(d) \times Y_{\text{col}(Z_W)}(d)$. The subgraph induced by fixing w_1 and w_2 in the first step is isomorphic to a graph that contains *all* rank $2m$ matrices in $\mathbb{F}_2^{n'}$ for some n' , that are a direct sum with some matrix Z_W (independent of Y). Thus every such subgraph is analyzed using the same tools and ideas as in the proof of Lemma 47.

In order to apply Lemma 16, we also need to show that the decomposition graph $\mathcal{H}_\tau$ expands (see its definition in Section 2.1.1). This is the bipartite graph where one side is the vertices in $\mathcal{L}_{Y,W}$, and the other side are the pairs of subspaces $Y_{\text{row}(Z_W)}(d) \times Y_{\text{col}(Z_W)}(d)$. A matrix $A \in \mathcal{L}_{Y,W}$ is connected to (w_1, w_2) if $\text{row}(A) \subseteq w_1$ and $\text{col}(A) \subseteq w_2$.

To analyze this graph, we reduce its expansion to the expansion of containment graphs in Y . We make the observation that we only care about $\text{row}(A)$ and $\text{col}(A)$, since if A and A' have the same rowspan and colspan, then they also have the same neighbors. To say this

in more detail, $\mathcal{H}_\tau \cong \mathcal{B} \otimes \mathcal{Q}$ where $\mathcal{Q}$ is some complete bipartite graph, and $\mathcal{B}$ replaces the matrices in $\mathcal{H}_\tau$ with $Y_{\text{row}(Z_W)}(d-2m) \times Y_{\text{col}(Z_W)}(d-2m)$ (here every matrix A is projected to $(\text{row}(A), \text{col}(A))$ in $\mathcal{B}$). The graph B' itself is a tensor product of the containment graph of $Y_{\text{row}(Z_W)}$ and $Y_{\text{col}(Z_W)}$. Fortunately, as we discussed in Section 2.3, this graph is an excellent expander, which concludes the proof.

Given these components we prove Theorem 42.

Proof of Theorem 42. Let us verify all parameters $[d', n', D', \lambda']$. The dimension d' is clear from the construction.

The size $|Y'(1)| \leq |Y(1)|^{2 \cdot \frac{d}{2}}$ since it is enough to specify d vectors to construct a rank $\frac{d}{2}$ matrix $M = \sum_{j=1}^{\frac{d}{2}} e_j \otimes f_j$ (of course we have also double counted since many decompositions correspond to the same matrix). Thus $\log |Y'(1)| \leq d \cdot D = D'$.

Let us verify that the ambient space is $n' = n^2$ dimensional, or in other words that $Y'(1)$ really spans all $n \times n$ matrices (recall that we assumed without loss of generality that Y 's ambient space is $\mathbb{F}_2^n$). For this it is enough to show that there are n^2 linearly independent matrices inside $Y'(1)$. Let $e_1, e_2, \dots, e_n \in Y(1)$ be n linearly independent vectors. The ambient space of Y is $\mathbb{F}_2^n$. Let us show that $e_i \otimes e_j$ is spanned by the matrices inside $Y'(1)$. Indeed, choose some subspaces $w_i, w_j \in Y(d)$ such that $e_i \in w_i, e_j \in w_j$. Then $Y'(1)$ contains all rank $\frac{d}{2}$ matrices in $\mathbb{F}_2^{n^2}$ whose row and column spans are inside w_i, w_j . These matrices span all matrices whose row and column spans are in w_i, w_j and in particular they span $e_i \otimes e_j$.

Finally let us show that all links are λ' -expanders. Fix $W \in Y'(\ell)$. By Lemma 46

$$Y'_W \cong (\mathcal{T}^m)^{\otimes 2^\ell - 1} \otimes \mathcal{L}_{Y,W}$$

so $\lambda(Y'_W) \leq \max\{\lambda(\mathcal{T}^m), \lambda(\mathcal{L}_{Y,W})\}$. Both are λ -expanders by Lemma 47 and Lemma 48 when the conditions of the theorem are met.

To sum up, Y' is a $[d', n', D', \lambda']$ -expanding poset. ◀

► **Remark 49.** For the readers who are familiar with [50], we remark that contrary to that paper, we do not use the strong trickling down machinery developed in [69]. [69] show that expansion in the top-dimensional links trickles down to the links of lower dimension (with a worse constant, but still sufficient for our theorem). A priori, one may think that using [69] could simplify the proof. However, in our poset all links have the same abstract structure, so we might as well analyze all dimensions at once. Moreover, one observes that in many cases the bound on the expansion of these posets actually *improves* for lower dimensions.

3.3 Proof of the decomposition lemma - Lemma 46

The proof of Lemma 46 uses the following claim proven in [50].

▷ **Claim 50** ([50, Proposition 55]). Let $Y^{all} = \mathcal{SP}(\mathcal{G})^{\leq d'}$ and let $W \in Y^{all}(\ell)$. Then there exists matrices $Z_1, Z_2, \dots, Z_{2^\ell - 1}$ of rank $\frac{d}{2^\ell}$, whose direct sum is denoted by $Z_W = \bigoplus_{i=1}^{2^\ell - 1} Z_i$, such that the link graph

$$Y_W^{all} \cong \left(\bigotimes_{j=1}^{2^\ell - 1} \mathcal{T}^{Z_j} \right) \otimes \mathcal{H}_{Z_W, m}.$$

This is not just an abstract isomorphism. The details are as follows. First, [50] observes that given $W = \text{Im}(\widehat{s})$, for some admissible s , the set $\text{Im}(s) := \{Z_j \mid \exists x \in \mathbb{F}_2^\ell \setminus \{0\}, s(x) = Z_j\}$ does not depend on s , only on U . We denote it by $Q_W = \{Z_j\}_{j=1}^{2^\ell - 1}$.

Intuitively, the following proposition follows from the 'Hadamard-like' construction of the Grassmann, and the intersection patterns of Hadamard code.

► **Proposition 51** ([50, Lemmas 59,61,62]). *Let $Y^{all} = \mathcal{SP}(\mathcal{G})^{\leq d'}$, let $W \in Y^{all}(\ell)$, let $Q_U = \{Z_1, Z_2, \dots, Z_{2^\ell-1}\}$ be as above, with $Z_W = \bigoplus_{j=1}^{2^\ell-1} Z_j$. For every matrix N such that $W + \text{span}(N) \in Y_U^{all}(\ell+1)$ there exists unique matrices $A_1, A_2, \dots, A_{2^\ell}$ such that*

1. $\text{rank}(A_j) = \frac{d}{2^{\ell+1}}$.
2. $N = \sum_{j=1}^{2^\ell} A_j$.
3. For $j = 1, 2, \dots, 2^\ell - 1$ the matrices $A_j \in \mathcal{T}^{Z_j}$. That is, $A_j \leq Z_j$.
4. The matrix $A_{2^\ell} \in \mathcal{H}_{Z_W, m}$. That is, $\text{row}(A_{2^\ell})$ (resp. $\text{col}(A_{2^\ell})$) intersects trivially with $\text{row}(Z_W)$ (resp. $\text{col}(Z_W)$).
5. The function $N \mapsto (A_1, A_2, \dots, A_{2^\ell})$ is an isomorphism

$$Y_W^{all} \cong \left(\bigotimes_{j=1}^{2^\ell-1} \mathcal{T}^{Z_j} \right) \otimes \mathcal{H}_{Z_W, m}$$

We restate and reprove this proposition in Section 7. The proof our decomposition lemma is short given the one from [50]. For this (and this only) we need to slightly alter Definition 40 and Definition 45 so that we take into account two posets instead of one. Let Y_1, Y_2 be two posets of the same dimension d .

► **Definition 52** (Two-poset admissible function, generalizing Definition 39). *An admissible function with respect to Y_1, Y_2 to be a function $s : \mathbb{F}_2^i \setminus \{0\} \rightarrow \mathbb{F}_2^{n \times n}$ as in Definition 39 with the distinction that the row and column spaces of the matrix $M_s = \bigoplus_{v \in \mathbb{F}_2^i \setminus \{0\}} s(v)$ are*

$$\text{row}(M_s) \in Y_1(2^{d'} - 2^{d'-i}), \text{col}(M_s) \in Y_2(2^{d'} - 2^{d'-i}).$$

► **Definition 53** (Two-poset matrix poset, generalizing Definition 40). *We denote by $\mathcal{SP}(Y_1, Y_2)$ the matrix poset defined in Definition 40 such that the admissible functions are with respect to Y_1 and Y_2 .*

► **Definition 54** (Two-poset sparsified link graph, generalizing Definition 45). *For $W \in \mathcal{SP}(Y_1, Y_2)$ we denote by $\mathcal{L}_{Y_1, Y_2, W}$ the graph in Definition 45, with the distinction that all sums of row spaces are required to be in Y_1 and all column spaces are required to be in Y_2 (instead of them all being in the same poset Y).*

Proof of Lemma 46. Fix W and its corresponding $Q_W = \{Z_1, Z_2, \dots, Z_{2^\ell-1}\}$ and Z_W . Conditioned on W , the first step of the sampling distribution in Definition 40 is the same as sampling $w_1 \supset \text{row}(Z_W)$ and $w_2 \supset \text{col}(Z_W)$ (where $w_1, w_2 \in Y(d)$). Fixing the pair w_1, w_2 and conditioning on all three W, w_1, w_2 , sampling an edge in the link Y_U' is the same as sampling an edge in the link of W in the complex $\mathcal{SP}(\mathcal{G}(w_1), \mathcal{G}(w_2))$. As both $\mathcal{G}(w_1)$ and $\mathcal{G}(w_2)$ are isomorphic, $\mathcal{SP}(\mathcal{G}(w_1), \mathcal{G}(w_2)) \cong \mathcal{SP}(\mathcal{G}(w_1), \mathcal{G}(w_1)) \cong \mathcal{SP}(\mathcal{G}(w_1))^{12}$. We are back to the complete Grassmann case and are allowed to use Proposition 51. Let $\mathcal{B}(W, w_1, w_2)$ be the subgraph obtained when fixing and conditioning on W, w_1, w_2 . By Proposition 51, we can decompose this graph to a tensor product

$$\mathcal{B}(W, w_1, w_2) \cong \left(\bigotimes_{j=1}^{2^\ell-1} \mathcal{T}^{Z_j} \right) \otimes \mathcal{L}_{\mathcal{G}(w_1), \mathcal{G}(w_2), W}.$$

¹² Recall that $\mathcal{G}(w_i)$ is the Grassmann poset containing all subspaces of w_i .

The first $2^\ell - 1$ components are independent of w_1 and w_2 ; they are constant among all conditionals (and are all isomorphic to $\mathcal{T}^m$). Hence this is isomorphic to $(\mathcal{T}^m)^{\otimes 2^\ell - 1} \otimes \mathcal{R}$ where $\mathcal{R}$ is the graph corresponding to a subgraph of $\mathcal{L}_{Y,W}$ obtained by conditioning the edge-sampling process of $\mathcal{L}_{Y,W}$ to sampling w_1 and w_2 in the first step. Thus the graph $\mathcal{R} = \mathcal{L}_{\mathcal{G}(w_1), \mathcal{G}(w_2), W}$ and the lemma is proven. $\blacktriangleleft$

3.4 Expansion of the sparsified link graph

In this section we use the decomposition lemma to reduce the analysis of the sparsified link graph to the case of the complete Grassmann and prove Lemma 48.

We use the following claim on the complete Grassmann.

$\triangleright$ **Claim 55.** Let $Y^{all} = \mathcal{SP}(\mathcal{G})^{\leq d'}$ be the poset of rank $\frac{d}{2}$ matrices for a sufficiently large d' . Let $W \in Y^{all}(\ell)$ and let $w_1, w_2 \in \mathcal{G}(d)$ be such that $\text{row}(Z_W) \subseteq w_1, \text{col}(Z_W) \subseteq w_2$. Then $\lambda(\mathcal{L}_{\mathcal{G}(w_1), \mathcal{G}(w_2), W}) \leq \gamma^m$ for some universal constant $\gamma \in (0, 1)$.

We prove this claim in the next section.

Proof of Lemma 48. Fix the subspace $W \in Y'(\ell)$ and recall the notation $w_R = \text{row}(Z_W), w_C = \text{col}(Z_W)$. We intend to use Lemma 15. The local decomposition we use is the following $\tau = (T, \mu_T)$. For every pair $(w_1, w_2) \in Y_{w_R}(d) \times Y_{w_C}(d)$ we define the graph $G_{(w_1, w_2)}$ which is the subgraph obtained by conditioning on (w_1, w_2) being sampled in the first step of Definition 45. The distribution μ_T is the distribution that samples $w_1, w_2 \in Y_W(d)$ independently. By definition of $\mathcal{L}_{Y,W}$ this is indeed a local-decomposition.

We observe that the subgraph of $\mathcal{L}_{Y,W}$ obtained by fixing w_1 and w_2 in the first step is just $G_{w_1, w_2} = \mathcal{L}_{\mathcal{G}(w_1), \mathcal{G}(w_2), W}$ and hence by Claim 55 it is a γ^m -expander.

We describe the local-to-global graph $\mathcal{H}_\tau$ in Lemma 15.

1. $T = Y_{w_R}(d) \times Y_{w_C}(d)$,
 2. $V = \{A \mid \text{rank}(A) = 2m, \text{row}(A) \oplus \text{row}(Z_W) \in Y, \text{col}(A) \oplus \text{col}(Z_W) \in Y(d - 2m)\}$,
- and is an edge between $(w_1, w_2) \in T$ and $A \in V$ if $\text{row}(A) \oplus \text{row}(Z_W) \subseteq w_1$ and $\text{col}(A) \oplus \text{col}(Z_W) \subseteq w_2$.

Observe that we only care about $\tau(A) = (\text{row}(A) \oplus \text{row}(Z_W), \text{col}(A) \oplus \text{col}(Z_W))$, and any two A_1, A_2 with $\tau(A_1) = \tau(A_2)$, have the same neighbor set. In other words, this graph is isomorphic to the bipartite tensor product $\mathcal{B}_R \otimes \mathcal{B}_C \otimes \mathcal{Q}_{1,k}$ where $\mathcal{B}_R$ is the containment graph between $Y_{w_R}(d)$ and $Y_{w_R}(d - 2m)$ (resp. $\mathcal{B}_C$ and w_C), and $\mathcal{Q}_{1,k}$ is the complete bipartite graph with k vertices on one side and a single vertex on the other side. The number k is the number of matrices A such that $(\text{row}(A) \oplus \text{row}(Z_W), \text{col}(A) \oplus \text{col}(Z_W)) = (u_1, u_2)$, for any pair of fixed subspaces u_1 and u_2 . By Claim 21 and the fact that Y is λ -expanding, the graph G is a $(0.61 + d\lambda)^{\frac{1}{2}(d-2m)} \leq \gamma^m$, where $\gamma < 1$ is some constant and the last inequality is because $\lambda \leq \frac{1}{4d}$.

By Lemma 15, $\lambda(\mathcal{L}_{Y,U}) \leq \lambda(G)^2 + O(\gamma^m) = \gamma^m$. $\blacktriangleleft$

4 Johnson complexes and local spectral expansion

In this section we construct a family of simplicial complexes which we call Johnson complexes whose degree is polynomial in their size. The underlying 1-skeletons of Johnson complexes are abelian Cayley graphs $\text{Cay}(\mathbb{F}_2^n, \binom{[n]}{\leq \varepsilon n})$. We prove in this section that these complexes are local spectral expanders and in particular the links of all faces except for the trivial face have spectral expansion strictly smaller than $\frac{1}{2}$. We then compare the Johnson complexes with the 2-dimensional local spectral expanders constructed from random geometric graphs [74] and

show that the former complexes can be viewed as a derandomization of the latter random complexes. On the way we also define Johnson Grassmann posets $\{Y_{\varepsilon,n}\}_n$, analogous to the matrix posets given in Section 3.1, and prove in Section 4.4 that Johnson complexes are exactly $X(\mathbb{F}_2^n, \beta(Y_{\varepsilon,n}))$, which are Cayley complexes whose vertex links are basifications of $Y_{\varepsilon,n}$ s (Definition 36).

► **Remark 56.** Later in Section 7, we shall define the induced Grassmann posets which generalize both the Johnson Grassmann posets and the matrix Grassmann posets, and we shall provide a unified approach to show local spectral expansion for induced Grassmann posets. However in this section, we will present a more direct proof of local expansion that is specific to Johnson complexes.

We name our construction Johnson complexes due to their connection to Johnson graphs. In particular, all the vertex links in a Johnson complex are generalization of Johnson graphs to simplicial complexes. So we start by giving the definition of Johnson graphs and then provide the construction of Johnson complexes.

► **Definition 57 (Johnson graph).** *Let $n > k > \ell$ be integers. The adjacency matrix Johnson graph $J(n, k, \ell)$ is the (uniformly weighted) graph whose vertices are all sets of size k inside $\{1, 2, \dots, n\}$ and two sets t_1, t_2 are connected if $|t_1 \cap t_2| = \ell$.*

Next we define Johnson complexes. We shall use $wt(x)$ to denote the Hamming weight of a binary string x .

► **Definition 58 (Johnson complex $X_{\varepsilon,n}$).** *Let k and be any integer. Let $n \in \mathbb{N}$ and let $\varepsilon \in (0, \frac{1}{2}]$ be such that $2^k \mid \varepsilon n$. Let $S_{\varepsilon,n} = \{x \in \mathbb{F}_2^n \mid wt(x) = \varepsilon n\}$ (abbreviate as S_ε when n is clear from context). The ε -weight Johnson complex $X = X_{\varepsilon,n}$ is the k -dimensional simplicial complex defined as follows.*

$$X(0) = \{x \in \mathbb{F}_2^n \mid wt(x) \text{ is even}\},$$

$$X(1) = \{\{x, x + s\} \mid x \in X(0), s \in S_\varepsilon\},$$

$$X(k) = \{\{x, x + s_1, \dots, x + s_k\} \mid x \in X(0), \forall i \in [k] s_i \in S_\varepsilon,$$

$$\forall \emptyset \subsetneq T \subseteq [k] \ wt\left(\sum_{i \in T} s_i\right) = \varepsilon n\}.$$

For the rest of this work, we fix k, n, ε as above.

It turns out that the weight constraints on k -faces in a Johnson complex gives rise to a Hadamard-like intersection pattern for elements in the link faces. Any $(k-1)$ -face in the link of x can be written as $\{x + s_1, \dots, x + s_k\}$. If we identify each s_i with the set of nonzero coordinates in it, these k sets always satisfy that any two distinct sets intersect at exactly $\varepsilon n/2$ coordinates, and in general any $\ell \leq k$ distinct s_i s intersect at $\varepsilon n/2^{i-1}$ coordinates. We note that this is similar to the intersection pattern of k distinct codewords in a Hadamard code.

The main result of this section is proving that all links of this complex are good expanders.

► **Lemma 59 (Local spectral expansion of Johnson complexes).** *For any $0 \leq m \leq k-2$, the link graph of any m -face in the k -dimensional Johnson complex $X_{\varepsilon,n}$ is a two-sided $(\frac{1}{2} - \frac{\varepsilon}{2^{m+1}})$ -spectral expander. In particular, $X_{\varepsilon,n}$ is a k -dimensional two-sided $(\frac{1}{2} - \frac{\varepsilon}{2^{k-1}})$ -local spectral expander.*

The key observation (Proposition 62) towards showing this lemma is that the link graphs of m -faces are the tensor product of Johnson graphs.

Once this decomposition lemma is proven, we can use the second eigenvalue bounds for Johnson graphs to show that the links are $(\frac{1}{2} - \frac{\varepsilon}{2^{m+1}})$ -spectral expanders. We note that this construction breaks the $\frac{1}{2}$ link expansion barrier of known elementary constructions of local spectral HDXs (discussed in the introduction). Because our construction breaks this barrier, using the Oppenheim's trickle down theorem [87], we directly get that the 1-skeleton of the Johnson complex is also a $(1 - 2\varepsilon)$ -spectral expander.

We now outline the rest of this section. First, we prove the link decomposition lemma Proposition 62. Then using this lemma we show that Johnson complexes are basifications of Johnson Grassmann posets. Afterwards, we apply this lemma to prove the main result. Finally, we conclude this section by comparing the density parameters of the Johnson complexes with those of spectrally expanding random geometric complexes.

4.1 Link decomposition

In this part we prove the link decomposition result. Note that by construction the links of all vertices $X_{\varepsilon,n}$ are isomorphic. So it suffices to focus on the link of $\mathbf{0} \in \mathbb{F}_2^n$ (denoted by $X_{\mathbf{0}}$) to show local spectral expansion for $X_{\varepsilon,n}$. The link decomposition is a consequence of the Hadamard-like intersection patterns of the faces in $X_{\mathbf{0}}$ as explained earlier.

For the rest of the section, we use $C_i \subset [n]$ to denote the set of nonzero coordinates in a vector x_i and use M to denote 2^m . We now state the face intersection pattern lemma.

► **Lemma 60.** *Let m be an integer. Given a set $\{x_1, \dots, x_m\} \subset \mathbb{F}_2^n$ and a subset $B \subset [m]$, define*

$$C_B = \bigcap_{i \in B} C_i \bigcap_{j \notin B} \overline{C_j} \text{ and } c_B = |C_B| .$$

Then $\{x_1, \dots, x_m\} \in X_{\mathbf{0}}(m-1)$ if and only if for any subset $B \subset [m]$,

$$c_B = \begin{cases} \frac{2\varepsilon n}{M} & B \neq \emptyset \\ n - \frac{M-1}{M} \cdot 2\varepsilon n & \text{otherwise} \end{cases} .$$

Proof. We first verify that every set $\{x_1, \dots, x_m\}$ satisfying the M cardinality constraints above is in $X_{\mathbf{0}}(m-1)$. Recall that by definition $\{x_1, \dots, x_m\} \in X_{\mathbf{0}}(m-1)$ if and only if they satisfy the $M-1$ weight equations

$$\forall S \subset [m], S \neq \emptyset, \text{ wt} \left(\sum_{i \in S} x_i \right) = \varepsilon n .$$

Observe that the set of nonzero coordinates in $\sum_{i \in S} x_i$ are precisely $j \in [n]$ that are nonzero in odd number of vectors in $\{x_i \mid i \in S\}$

$$\text{Set} \left(\sum_{i \in S} x_i \right) = \bigoplus_{i \in S} C_i = \bigcup_{\substack{B \subseteq [m], \\ |B \cap S| \text{ odd}}} C_B ,$$

and that by definition the set $\{C_B\}_{B \subseteq [m]}$ is a partition of $[n]$. Therefore the weights can be rewritten as

$$\forall S \subset [m], S \neq \emptyset, \text{ wt} \left(\sum_{i \in S} x_i \right) = \sum_{\substack{B \subseteq [m], \\ |B \cap S| \text{ odd}}} c_B . \quad (4.1)$$

Furthermore, for any nonempty subset S there are $M/2$ nonempty sets $B \subset [m]$ whose intersection with S has odd cardinality: let $t = |S|$, then the number of S 's subsets of odd cardinality is

$$\sum_{i \text{ odd}} \binom{t}{i} = \frac{1}{2} \left(\sum_{i \leq t} \binom{t}{i} - \left(\sum_{i \text{ even}} \binom{t}{i} - \sum_{i \text{ odd}} \binom{t}{i} \right) \right) = \frac{1}{2} ((1+1)^t - (1-1)^t) = 2^{t-1}.$$

So the total number of such B is $2^{t-1} \cdot 2^{m-t} = M/2$. Therefore if $c_B = \frac{2\varepsilon n}{M}$ when $B \neq \emptyset$, then $wt(\sum_{i \in S} x_i) = M/2 \cdot 2\varepsilon n/M = \varepsilon n$. Thus under this condition $\{x_1, \dots, x_m\} \in X_{\mathbf{0}}(m-1)$.

Next we prove by induction that any $\{x_1, \dots, x_m\} \in X_{\mathbf{0}}(m-1)$ satisfies the conditions on c_B . The statement trivially holds when $m = 0$. Suppose the same statement holds for all i -faces where $i \leq m-2$.

By (4.1), we can deduce that every $\{x_1, \dots, x_{m-1}\} \in X_{\mathbf{0}}(m-2)$ satisfies:

$$\forall S \subset [m-1], S \neq \emptyset, \quad \sum_{B: |B \cap S| \text{ odd}} c_B = \varepsilon n. \quad (4.2)$$

Then by the induction hypothesis we know that $c_B = \frac{4\varepsilon n}{M}$ for $B \neq \emptyset$ is the unique solution to the set of constraints above.

Now consider the set of constraints for the $(m-1)$ -face $\{x_1, \dots, x_m\}$. There are in total M variables and we denote them $\{c'_B, c'_{B \cup \{m\}}\}_{B \subset [m-1]}$ where $c'_B = |C_B \cap \overline{C_m}|$ and $c'_{B \cup \{m\}} = |C_B \cap C_m|$. The new variables are related to variables in $\{c_B\}_{B \subset [m-1]}$ via the identity $c_B = c'_B + c'_{B \cup \{m\}}$. So (4.2) can be expanded as:

$$\forall S \subset [m-1], S \neq \emptyset, \quad \sum_{B: |B \cap S| \text{ odd}} c_B = \sum_{B: |B \cap S| \text{ odd}} c'_B + \sum_{B: |B \cap S| \text{ odd}} c'_{B \cup \{m\}} = \varepsilon n. \quad (4.3)$$

Additionally, via (4.1) the weight constraints $wt(x_m + \sum_{i \in S} x_i) = \varepsilon n$ can be rewritten as:

$$\forall S \subset [m-1], S \neq \emptyset, \quad \sum_{B: |B \cap S| \text{ odd}} c'_B + \sum_{B: |B \cap S| \text{ even}} c'_{B \cup \{m\}} = \varepsilon n. \quad (4.4)$$

Finally using the constraint $wt(x_m) = \varepsilon n$, we obtain equations:

$$\forall S \subset [m-1], S \neq \emptyset, \quad \sum_{B: |B \cap S| \text{ odd}} c'_{B \cup \{m\}} + \sum_{B: |B \cap S| \text{ even}} c'_{B \cup \{m\}} = |C_m| = \varepsilon n. \quad (4.5)$$

From (4.3), (4.4), and (4.5) we can deduce that

$$\forall S \subset [m-1], S \neq \emptyset, \quad \sum_{B: |B \cap S| \text{ odd}} c'_{B \cup \{m\}} = \varepsilon n/2. \quad (4.6)$$

The $M/2 - 1$ constraints in (4.6) are exactly the same as the set of constraints in (4.2) except that the RHS values are $\varepsilon n/2$ instead of εn . Therefore by the induction hypothesis, the unique solution is $\forall B \neq \emptyset, c'_{B \cup \{m\}} = \frac{2\varepsilon n}{M}$. As a result $c'_B = c_B - c'_{B \cup \{m\}} = \frac{2\varepsilon n}{M}$. Finally we also have $c'_m = \varepsilon n - (M/2 - 1) \cdot \frac{2\varepsilon n}{M} = \frac{2\varepsilon n}{M}$ and $c'_\emptyset = c_\emptyset - c'_m = n - \frac{M-1}{M} \cdot 2\varepsilon n$. Thus we complete the induction step. ◀

► **Corollary 61.** Let m be an integer and $M = 2^m$. For any $\{x_1, \dots, x_m\} \in X_{\mathbf{0}}(m-1)$, x_m 's nonzero coordinate set C_m can be decomposed into the disjoint union of $M/2$ sets each of size $2\varepsilon n/M$ as follows,

$$C_m = \cup_{B \subset [m-1]} C_{B \cup \{m\}}, \text{ where } C_{B \cup \{m\}} = \bigcap_{i \in B \cup \{m\}} C_i \bigcap_{j \notin B \cup \{m\}} \overline{C_j}.$$

Applying this coordinate decomposition to vertices and edges in the link graph of any face in X_0 , we obtain the following characterization of the graphs.

► **Proposition 62.** *Let $0 \leq m \leq k-2$ be an integer and let $M = 2^m$. For any m -face in $X_{\varepsilon,n}$, the underlying graph of its link is isomorphic to $J(\frac{2\varepsilon}{M}n, \frac{\varepsilon}{M}n, \frac{\varepsilon}{2M}n)^{M-1} \otimes J(cn, \frac{\varepsilon}{M}n, \frac{\varepsilon}{2M}n)$ where $c = 1 - \frac{M-1}{M} \cdot 2\varepsilon$.*

Proof. Give an m -face $\{x, x+x_1, \dots, x+x_m\}$ in $X_{\varepsilon,n}$, we note that by construction its link is isomorphic to the link of $\{0, x_1, \dots, x_m\}$. So without loss of generality we only consider m -faces of the form above and we denote such a face by t . The by definition the vertex set in the link of t can be written as

$$X_t(0) = \left\{ x_{m+1} \mid \forall \emptyset \subsetneq S \subset [m+1], \text{ wt} \left(\sum_{i \in S} x_i \right) = \varepsilon n \right\}.$$

By the decomposition lemma Corollary 61, we have the following isomorphism between the link vertices and collections of disjoint sets in $\binom{n}{\varepsilon n/M}$. First recall the notation $C_B = \bigcap_{i \in B} C_i \cap_{j \in [m] \setminus B} \overline{C_j}$. Then the isomorphism $\phi : X_t(0) \rightarrow \prod_{B \subseteq [m]} \binom{C_B}{\varepsilon n/M}$ is given by

$$\phi(x_{m+1}) = \{C_{m+1} \cap C_B\}_{B \subseteq [m]}.$$

We use $\phi(x_{m+1})[B]$ to denote the set $C_{m+1} \cap C_B$.

Similarly, using Corollary 61 we can write the edge set as:

$$X_t(1) = \left\{ \{x_{m+1}, x'_{m+1}\} \mid x_{m+1}, x'_{m+1} \in X_0(0), \right. \\ \left. \forall B \subset [m], |\phi(x_{m+1})[B] \cap \phi(x'_{m+1})[B]| = \frac{\varepsilon n}{2M} \right\}.$$

Therefore the adjacency matrix of t 's link graph has entries:

$$A_t[x_{m+1}, x'_{m+1}] = \prod_{B \subseteq [m]} \mathbf{1} \left[|\phi(x_{m+1})[B] \cap \phi(x'_{m+1})[B]| = \frac{\varepsilon n}{2M} \right].$$

The product structure of A_t gives rise to a natural decomposition into tensor products of matrices $H_B \in \{0, 1\}^{\binom{C_B}{\varepsilon n/M} \times \binom{C_B}{\varepsilon n/M}}$ where

$$\forall U, V \in \binom{C_B}{\varepsilon n/M}, H_B[U, V] = \mathbf{1} \left[|U \cap V| = \frac{\varepsilon n}{2M} \right].$$

Furthermore since every nonempty $B \subset [m]$ has $|C_B| = c_B = \frac{2\varepsilon}{M}n$ (Lemma 60), H_B is the adjacency matrix of the Johnson graph $J(\frac{2\varepsilon}{M}n, \frac{\varepsilon}{M}n, \frac{\varepsilon}{2M}n)$. In the case $B = \emptyset$, $|C_B| = cn$ (Lemma 60), and so $H_\emptyset$ is the adjacency matrix of $J(cn, \frac{\varepsilon}{M}n, \frac{\varepsilon}{2M}n)$. In conclusion, the link graph of t is isomorphic to the tensor product $J(\frac{2\varepsilon}{M}n, \frac{\varepsilon}{M}n, \frac{\varepsilon}{2M}n)^{M-1} \otimes J(cn, \frac{\varepsilon}{M}n, \frac{\varepsilon}{2M}n)$. ◀

4.2 Local spectral expansion

We now finish the local spectral expansion proof for Johnson complexes using Proposition 62 together with the following Johnson graph eigenvalue bounds and the trickle-down theorem.

► **Theorem 63** ([24]). *Let A be the unnormalized adjacency matrix of $J(n, k, \ell)$. The eigenvalues of A are $\lambda_0, \dots, \lambda_k$ where*

$$\lambda_t = \sum_{i=0}^t (-1)^{t-i} \binom{k-i}{\ell-i} \binom{n-k+i-t}{k-\ell+i-t} \binom{t}{i}.$$

We note that in the theorem above λ_i are unnormalized eigenvalues of the Johnson graphs. In the rest of the paper when we talk about spectral expansion, we consider normalized eigenvalues whose values are always in $[-1, 1]$.

We specifically care about the second eigenvalues of the Johnson graphs with $n \geq 2k$. The case of $n = 2k$ is proved in [70].

► **Lemma 64** ([70]). *The Johnson graph $J(2k, k, k/2)$'s eigenvalues satisfy $\max_{t=1, \dots, k} \frac{|\lambda_t|}{\lambda_0} \leq \frac{|\lambda_2|}{\lambda_0} = \frac{1}{2k-2}$.*

In other words, the graph $J(2k, k, k/2)$ is a two-sided $\frac{1}{2k-2}$ -spectral expander.

For the case $n > (2 + \eta)k$, we show the following lemma (proof in Section E):

► **Lemma 65.** *For any constant $\eta > 0$, the Johnson graph $J((2 + \eta)k, k, k/2)$'s eigenvalues satisfy that $\max_{t=1, \dots, k} \frac{|\lambda_t|}{\lambda_0} \leq \frac{\eta}{2(1+\eta)}$. In other words the graph is a $\frac{\eta}{2(1+\eta)}$ -spectral expander.*

The trickle-down theorem states that local spectral expansion implies expansion of all links.

► **Theorem 66** ([87]). *Let $\lambda \geq 0$. Let X be a 2-dimensional connected simplicial complex and assume that for any vertex $v \in X(0)$, the link X_v is a two-sided λ -spectral expander, and that the 1-skeleton of X is connected. Then the 1-skeleton of X is a two-sided $\frac{\lambda}{1-\lambda}$ -spectral expander. In particular, X is a two-sided $\frac{\lambda}{1-\lambda}$ local spectral expander.*

Now we are ready to prove the main result of this section.

Proof of Lemma 59. Let t be any m -face in the complex. By Proposition 62, the link graph of t is $G_t = J(\frac{2\varepsilon}{M}n, \frac{\varepsilon}{M}n, \frac{\varepsilon}{2M}n)^{M-1} \otimes J(cn, \frac{\varepsilon}{M}n, \frac{\varepsilon}{2M}n)$. Since eigenvalues of $A_1 \otimes A_2$ are $\{\lambda_1 \cdot \lambda_2 \mid \lambda_1 \in \text{Spec}(A_1), \lambda_2 \in \text{Spec}(A_2)\}$, by Equation (2.1) the second largest eigenvalues satisfy $\lambda(A_1 \otimes A_2) = \max(\lambda(A_1), \lambda(A_2))$. Therefore the link graph G_t of t has $|\lambda|_2(G_t) \leq \max\left(\frac{1}{2\varepsilon n/M-2}, \frac{1-2\varepsilon}{2(1-(2-M^{-1})\varepsilon)}\right) = \frac{1}{2} \cdot \frac{1-2\varepsilon}{1-(2-M^{-1})\varepsilon} \leq \frac{1}{2} \cdot \left(1 - \frac{\varepsilon}{M}\right)$ (via Lemma 64 and Lemma 65). Therefore the k -dimensional Johnson complex $X_{\varepsilon, n}$ is a two-sided $\left(\frac{1}{2} - \frac{\varepsilon}{2^{k-1}}\right)$ -local spectral expander.

Furthermore, applying the trickle-down theorem to the 2-skeleton $X_{\varepsilon, n}^{\leq 2}$ we get that the underlying graph of the complex is a two-sided $\frac{(1-2\varepsilon)/(2-2\varepsilon)}{1-(1-2\varepsilon)/(2-2\varepsilon)} = (1-2\varepsilon)$ -spectral expander. ◀

4.3 Comparison to local spectral expanders from random geometric complexes

In this part, we compare the structure of the Johnson complexes to that of the random geometric complexes from [74]. We observe that the former can be viewed as a derandomization of the latter.

The random 2-dimensional geometric complex (RGC) $\text{Geo}_d(N, p)$ is defined by sampling N i.i.d. points $\{x_i\}_{i \in [N]}$ from the uniform distribution over the d -dimensional unit sphere, adding an edge between ij if $\langle x_i, x_j \rangle \geq \tau$ where τ is picked so that the marginal probability of an edge is p , and adding a 2-face if the three points are pairwise connected. In the following parameter regime the complexes have expanding vertex links.

► **Theorem 67** (Rephrase of Theorem 1.6 in [74]). *For every $\delta < \frac{1}{2}$, there exists C_δ and $c = \Theta\left(\frac{1}{\log(1/\delta)}\right)$ such that when $H \sim \text{Geo}_d(N, N^{-1+c})$ for $d = C_\delta \log N$, with high probability every vertex link of H is a two-sided $(1 - 2\delta)$ -local spectral expander whose vertex links are $\left(\frac{1}{2} - \delta\right)$ -spectral expanders.*

The Johnson complex $X := X_{\varepsilon,n}$ with $N = |X(0)|$ can be viewed as an approximate derandomization of the RGCs in $\mathbb{R}^n$ over N vertices. Note that the points in $X(0)$ (which are boolean vectors) embed in the natural way to $\mathbb{R}^n$. Furthermore, these points in $\mathbb{R}^n$ all lie on a sphere of radius $\frac{\sqrt{n}}{2}$ centered at $[\frac{1}{2}, \dots, \frac{1}{2}]$. Two points are connected in X if the L_2 distance between their images in $\mathbb{R}^n$ is exactly $\sqrt{\varepsilon n}$. The distinction from edges in RGCs is that in Johnson complexes points of distance $< \sqrt{\varepsilon n}$ are not connected. Albeit the analysis in [74] shows that most neighbors of a vertex have distance $\approx \sqrt{\varepsilon n}$ so this is not a significant distinction. Similar to RGCs, three points in X form a 2-face if they are pairwise connected. By Lemma 59 the resulting 2-dimensional Johnson complex X has degree $N^{h(\varepsilon)}$ ¹³ and its vertex links are $(\frac{1}{2} - \frac{\varepsilon}{2})$ -spectral expanders.

► **Remark 68.** We further observe that while both constructions yield local spectral expanders with arbitrarily small polynomial vertex degree, the link graphs of the Johnson complexes are much denser. Use N to denote the number of vertices in the complexes. Then in the Johnson complexes the link graphs have vertex degree $\text{poly}(N)$ while in the RGCs the link graphs have average vertex degree $\text{polylog}(N)$. We remark that one can use a graph sparsification result due to [21] to randomly subsample the 2-face in $X_{\varepsilon,n}$ while preserving their link expansion. As a result we obtain a family of random subcomplexes of the Johnson complexes that are local spectral expanders with similar vertex degree and link degree as the random geometric complexes. Detailed descriptions and proofs are given in Section F.

4.4 Johnson Grassmann posets

In this part, we take a small digression to define the Johnson Grassmann posets $Y_{\varepsilon,n}$ and show that their basifications are exactly Johnson complexes $X_{\varepsilon,n}$.

► **Definition 69** (Johnson Grassmann posets $Y_{\varepsilon,n}$). *Let $d = \log 2\varepsilon n$. For every $i \in [d]$, define the admissible function set as*

$$\mathcal{S}_i = \left\{ s : \mathbb{F}_2^i \setminus \{0\} \rightarrow \mathbb{F}_2^n \mid \forall v \in \mathbb{F}_2^i \text{ } wt(s(v)) = 2^{d-i}, \text{ } wt\left(\sum_{v \in \mathbb{F}_2^i \setminus \{0\}} s(v)\right) = 2^d - 2^{d-i} \right\}.$$

Define the Hadamard encoding $\widehat{s} : \mathbb{F}_2^i \rightarrow \mathbb{F}_2^n$ of s as

$$\widehat{s}(y) = \sum_{v \in \mathbb{F}_2^i \setminus \{0\}} \langle y, v \rangle s(v).$$

For each admissible function $s \in \mathcal{S}_i$, use $\mathcal{G}_s$ to denote the Grassmann poset $\{Im(\widehat{s}|_U) \mid U \subseteq \mathbb{F}_2^i, U \text{ is a subspace}\}$

Then the Johnson Grassmann poset is $Y_{\varepsilon,n} = \bigcup_{s \in \mathcal{S}_d} \mathcal{G}_s$.

We first verify that these posets are well-defined.

▷ **Claim 70.** For all $1 \leq i \leq d$, $Y_{\varepsilon,n}(i) = \{Im(\widehat{s}) \mid s \in \mathcal{S}_i\}$.

Proof. The claim trivially holds for $i = d$. For $i < d$, we need to prove that (1) for every $t \in \mathcal{S}_i$, there exist $s \in \mathcal{S}_d$ and i -dimensional subspace $U \subseteq \mathbb{F}_2^d$ such that $Im(\widehat{t}) = Im(\widehat{s}|_U)$ and (2) for every $s \in \mathcal{S}_d$, any subspace $W \subset Im(\widehat{s})$ is in $Y_{\varepsilon,n}$.

¹³ h is the binary entropy function $h(p) = p \cdot \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$.

We first prove (1). Given any admissible function $t \in \mathcal{S}_i$, for every $v \in \mathbb{F}_2^i \setminus \{0\}$ and $u \in \mathbb{F}_2^{d-i}$, use $e_{t(v),u} \in \mathbb{F}_2^n$ to denote the vector that is nonzero only at the u -th nonzero coordinate of $t(v)$. Here we identify the vector u in a natural way with a number in $[2^{d-i}]$. Use z to denote the sum $z = \sum_{v \in \mathbb{F}_2^i \setminus \{0\}} t(v)$. Set U to be the subspace that spans the first i coordinates in $\mathbb{F}_2^d$, and set $s : \mathbb{F}_2^d \setminus \{0\} \rightarrow \mathbb{F}_2^n$ as follows

$$\forall v \in \mathbb{F}_2^i, u \in \mathbb{F}_2^{d-i}, s(v, u) = \begin{cases} e_{t(v),u} & v \neq 0, \\ e_{1-z,u} & \text{o.w.} \end{cases}$$

Then it is straightforward to verify that $s \in \mathcal{S}_d$ and that $\text{Im}(\hat{t}) = \text{Im}(\hat{s}|_U)$.

Next we check (2). Note that the codespace $\text{Im}(\hat{s})$ is a d -dimensional subspace in $\mathbb{F}_2^n$ whose nonzero elements have weight $2^d/2 = \varepsilon n$. Furthermore, consider any linear subspace $W \subset \text{Im}(\hat{s})$ of dimension $j < d$. Since $\hat{s}$ is an isomorphism, there exists a j -dimensional linear subspace $U \subset \mathbb{F}_2^d$ such that $W = \text{Im}(\hat{s}|_U)$ which is the image of $\hat{s}$ restricted to the domain U . Let $\psi \in \mathbb{F}_2^{d \times j}$ be a linear map from $\mathbb{F}_2^j$ to U , and $\phi \in \mathbb{F}_2^{j \times d}$ be such that $\phi\psi = \text{Id}$. Define the function $s_U : \mathbb{F}_2^j \setminus \{0\} \rightarrow \mathbb{F}_2^n$ as follows

$$s_U(v) = \sum_{v' \in U^\perp} s(\psi v + v').$$

By construction, s_U is admissible and furthermore for every $y \in \mathbb{F}_2^j$

$$\begin{aligned} \hat{s}_U(y) &= \sum_{v \in \mathbb{F}_2^j \setminus \{0\}} \langle y, v \rangle \cdot \sum_{v' \in U^\perp} s(\psi v + v') \\ &= \sum_{v'' \in \mathbb{F}_2^d \setminus \{0\}} \langle y, \phi v'' \rangle s(v'') = \sum_{v'' \in \mathbb{F}_2^d \setminus \{0\}} \langle \phi^\top y, v'' \rangle s(v'') = \hat{s}(\phi^\top y). \end{aligned}$$

We derive that $\text{Im}(\hat{s}_U) = \text{Im}(\hat{s}|_U) = W$. Thus $W \in Y_{\varepsilon,n}(j)$ and so the claim is proven. $\triangleleft$

► **Remark 71.** We observe that $Y_{\varepsilon,n}$ is a subposet of the complete Grassmann poset $\mathcal{G}(\mathbb{F}_2^n)$, and is obtained by applying the sparsification operator $\mathcal{SP}$ to $\mathcal{G}(\mathbb{F}_2^n)^{\leq \varepsilon n}$.

We now prove that $X_{\varepsilon,n}$ is the basification of $Y_{\varepsilon,n}$ using the link decomposition result.

► **Claim 72.** The k -dimensional Johnson complex $X_{\varepsilon,n}$ is the k -skeleton of the basification of $Y_{\varepsilon,n}$.

Proof. To prove the claim it suffices to show that any $\text{span}(x_1, x_2, \dots, x_k) \in Y_{\varepsilon,n}(k)$ if and only if all non-empty subset $T \subseteq [k]$ has $\text{wt}(\sum_{i \in T} x_i) = \varepsilon n$. In this proof we abuse the notation a bit and identify a vector in $\mathbb{F}_2^n$ with the set of its nonzero coordinates.

For the if direction, given $\{x_1, x_2, \dots, x_k\}$, construct $s : \mathbb{F}_2^k \setminus \{0\} \rightarrow \mathbb{F}_2^n$ such that $s(v) = \cap_{i=1}^k x_i^{v_i}$ where $x_i^{v_i}$ is the set of nonzero coordinates of x_i if $v_i = 1$, and the set of zero coordinates of x_i if $v_i = 0$. By Lemma 60, all such $s(v)$'s have weights exactly $\varepsilon n/2^{k-1}$. By construction their nonzero coordinates are disjoint from each other's. So s is an admissible function. Furthermore, let $\{e_1, e_2, \dots, e_k\}$ be elementary vectors in $\mathbb{F}_2^k$,

$$\begin{aligned} \hat{s}(e_i) &= \bigcup_{v \in \mathbb{F}_2^k \setminus \{0\}} \langle e_i, v \rangle s(v) \\ &= \bigcup_{v, \text{ s.t. } v_i=1} x_i \cap_{j \neq i} x_j^{v_j} \end{aligned}$$

$$\begin{aligned}
&= x_i \cap \bigcup_{v_{-i} \in \mathbb{F}_2^{k-1}} \cap_{j \neq i} x_j^{v_j} \\
&= x_i \cap [n] = x_i
\end{aligned}$$

Therefore $\text{span}(x_1, x_2, \dots, x_k) = \text{Im}(\widehat{s}) \in Y_{\varepsilon, n}(k)$.

For the only if direction, we can easily check that any $\text{Im}(\widehat{s}) \in Y_{\varepsilon, n}(k)$ contains only vectors of weight εn . So given any basis $\{x_1, x_2, \dots, x_k\}$ of $\text{Im}(\widehat{s})$, their nonzero linear combinations all have weight εn . Now the proof is complete. $\triangleleft$

5 Coboundary expansion and a van Kampen lemma

In this section we give the necessary preliminaries on coboundary expansion and also introduce a version of the famous van Kampen Lemma [97] that shall be used to prove coboundary expansion in the next section.

5.1 Preliminaries on coboundary expansion

We define coboundary expansion for general groups, closely following the definitions and notation in [28]. For a more thorough discussion on the connection between coboundary expansion and topology, we refer the reader to [28]. In the introduction we defined coboundary expansion as a notion of local testability of a code (which was a subspace of the functions on the edges). Instead of defining coboundary expansion for only $\mathbb{F}_2$ -valued functions, we give the most general version of coboundary expansion, which uses arbitrary groups.

Let X be a d -dimensional simplicial complex for $d \geq 2$ and let Γ be any group. Let $\vec{X}(1)$ be the directed edges in X and $\vec{X}(2)$ be the directed 2-faces. For $i = -1, 0$ let $C^i(X, \Gamma) = \{f : X(i) \rightarrow \Gamma\}$. We sometimes identify $C^{-1}(X, \Gamma) \cong \Gamma$. For $i = 1, 2$ let

$$C^1(X, \Gamma) = \left\{ f : \vec{X}(1) \rightarrow \Gamma \mid f(u, v) = f(v, u)^{-1} \right\}$$

and

$$\begin{aligned}
C^2(X, \Gamma) = \{ f : \vec{X}(2) \rightarrow \Gamma \mid \forall \pi \in \text{Sym}(3), (v_0, v_1, v_2) \in \vec{X}(2), \\
f(v_{\pi(0)}, v_{\pi(1)}, v_{\pi(2)}) = f(v_0, v_1, v_2)^{\text{sign}(\pi)} \}.
\end{aligned}$$

be the spaces of so-called *anti symmetric* functions on edges and triangles. For $i = -1, 0, 1$ we define operators $\delta_i : C^i(X, \Gamma) \rightarrow C^{i+1}(X, \Gamma)$ by

1. $\delta_{-1} : C^{-1}(X, \Gamma) \rightarrow C^0(X, \Gamma)$ is $\delta_{-1}h(v) = h(\emptyset)$.
2. $\delta_0 : C^0(X, \Gamma) \rightarrow C^1(X, \Gamma)$ is $\delta_0h(vu) = h(v)h(u)^{-1}$.
3. $\delta_1 : C^1(X, \Gamma) \rightarrow C^2(X, \Gamma)$ is $\delta_1h(vuw) = h(vu)h(uw)h(wv)$.

Let $\text{Id} = \text{Id}_i \in C^i(X, \Gamma)$ be the function that always outputs the identity element. It is easy to check that $\delta_{i+1} \circ \delta_i \equiv \text{Id}_{i+2}$ for all $i = -1, 0$ and $h \in C^i(X, \Gamma)$. Thus we denote by

$$Z^i(X, \Gamma) = \ker \delta_i \subseteq C^i(X, \Gamma),^{14}$$

$$B^i(X, \Gamma) = \text{Im} \delta_{i-1} \subseteq C^i(X, \Gamma),$$

and have that $B^i(X, \Gamma) \subseteq Z^i(X, \Gamma)$.

Henceforth, when the dimension i of the function f is clear from the context we denote $\delta_i f$ by δf .

Coboundary expansion is a property testing notion so for this we need a notion of distance. Let $f_1, f_2 \in C^i(X, \Gamma)$. Then

$$\text{dist}(f_1, f_2) = \mathbb{P}_{s \in \vec{X}(i)} [f_1(s) \neq f_2(s)]. \quad (5.1)$$

We also denote the weight of the function $\text{wt}(f) = \text{dist}(f, \text{Id})$.

► **Definition 73** (1-dimensional coboundary expander). *Let X be a d -dimensional simplicial complex for $d \geq 2$. Let $\beta > 0$. We say that X is a 1-dimensional β -coboundary expander if for every group Γ , and every $f \in C^1(X, \Gamma)$ there exists some $g \in C^0(X, \Gamma)$ such that*

$$\beta \text{dist}(f, \delta g) \leq \text{wt}(\delta f). \quad (5.2)$$

In this case we denote $h^1(X) \geq \beta$.

This definition in particular implies that $B^1 = Z^1$ for any coefficient group.

Coboundary expansion generalizes edge expansion

We note that 1-dimensional coboundary expanders are generalizations of edge expanders to simplicial complexes of dimension ≥ 2 . A graph $G = (V, E)$ is a 1-dimensional simplicial complex. For $\Gamma = \mathbb{F}_2$ we can analogously define function spaces $C^{-1}(G, \mathbb{F}_2)$, $C^0(G, \mathbb{F}_2)$, $C^1(G, \mathbb{F}_2)$, and operators δ_{-1} and δ_0 over G .

Observe that $B^0(G, \mathbb{F}_2)$ consists of constant functions over the vertices. Using the notations above we can write the definition of a 0-dimensional coboundary expander as follows.

► **Definition 74** (0-dimensional coboundary expander in $\mathbb{F}_2$). *A 1-dimensional simplicial complex $G = (V, E)$ is a 0-dimensional β -coboundary expander in $\mathbb{F}_2$ if for every $f \in C^0(G, \mathbb{F}_2)$ there exists some $g \in C^{-1}(G, \mathbb{F}_2)$ such that*

$$\beta \text{dist}(f, \delta_{-1}g) \leq \text{wt}(\delta_0 f). \quad (5.3)$$

Let us check that this is equivalent to the definition of β -edge expanders in terms of graph conductance. Viewing every function f as a partition of V into two sets V_f and $\bar{V}_f$ and viewing $\delta_0 f$ as the indicator function of whether an edge is in the cut between V_f and $\bar{V}_f$, we derive that

$$\min_{g \in C^{-1}(G, \mathbb{F}_2)} \text{dist}(f, \delta_{-1}g) = \min \left(\frac{|V_f|}{|V|}, \frac{|\bar{V}_f|}{|V|} \right), \quad \text{wt}(\delta_0 f) = \frac{|E(V_f, \bar{V}_f)|}{|E|}.$$

Then the condition Equation (5.3) is equivalent to the statement that for every vertex partition $(V_f, \bar{V}_f)$ of the graph $G = (V, E)$,

$$\beta \min \left(\frac{|V_f|}{|V|}, \frac{|\bar{V}_f|}{|V|} \right) \leq \frac{|E(V_f, \bar{V}_f)|}{|E|}.$$

This is the standard definition of a β -edge expander.

Coboundary Expansion and Local Testability

One should view coboundary expansion as a property testing notion. We call $B^1(X, \Gamma) \subseteq C^1(X, \Gamma)$ a code (as a subset, no linearity is assumed in the general case). Given a function in the ambient space $f \in C^1(X, \Gamma)$ one cares about its Hamming distance to the code, that

is, how many edges of f need to be changed in order to get a word in $B^1(X, \Gamma)$. Denote this quantity $\text{dist}(f, B^1(X, \Gamma))$. Coboundary expansion relates this quantity to another quantity, the fraction of triangles $\{u, v, w\}$ such that $\delta f(uvw) \neq Id$, namely $\text{wt}(\delta f)$.

We note that if $f = \delta g \in B^1(X, \Gamma)$, namely f satisfies the property of being in the function space, then one can verify that $\delta \delta g = Id$, or in other words, that $\text{wt}(\delta f) = 0$. In a β -coboundary expander we get a robust inverse of this fact. That is, that if f is ε -far from any function in $B^1(X, \Gamma)$, then its coboundary is not equal to the identity on at least a $\beta\varepsilon$ -fraction of the triangles.

This suggests the following natural test for this property.

1. Sample a random triangle $\{u, v, w\} \in X(2)$.
2. Accept if and only if $f(uv) = f(uw) \cdot f(vw)$ (or equivalently $\delta f(uvw) = Id$).

Note that $\text{wt}(\delta f(uvw))$ is the probability that f fails the test. Saying that a complex is a coboundary expander is saying that this test is sound, namely X is a β -coboundary expander if and only if for every $f : \vec{X}(1) \rightarrow \Gamma$

$$\text{dist}(f, B^1(X, \Gamma)) \leq \frac{1}{\beta} \mathbb{P}_{uvw} [f \text{ fails the test}]^{15} \quad (5.4)$$

Over $\mathbb{F}_2$, this view of coboundary expansion was first observed by [61], who used it for testing whether matrices over $\mathbb{F}_2$ have low rank. This view also appeared in [41] for a topological property testing notion called cover-stability. In later works, this test was as a component in the analysis of agreement tests [51, 27, 10]. We also refer the reader to [10, 29] for additional connections of coboundary expansion and local testing in the context of unique games problems and to [18] for its connection to group (te)stability.

5.2 Contractions, cycles and cones

One general approach to proving coboundary expansion is the cone method invented by [52] and used in many subsequent works to bound coboundary expansion of various complexes [77, 63, 71, 68, 28]. The following adaptation of the cone method to functions over non-abelian groups is due to [29]. We adopt their definition of cones, and after presenting the definition, we will explain the motivation behind it as well as its connection to isoperimetric inequalities of the complex.

We use ordered sequences of vertices to denote paths in X , and $\circ$ to denote the concatenation operator¹⁶, and for any path P , we use P^{-1} to denote the reverse path. Fixing a simplicial complex X , we define two symmetric relations on paths in X .

(BT) We say that $P_0 \stackrel{(BT)}{\sim} P_1$ if $P_i = Q \circ (u, v, u) \circ Q'$ and $P_{1-i} = Q \circ (u) \circ Q'$ where $i \in \{0, 1\}$.

(TR) We say that $P_0 \stackrel{(TR)}{\sim} P_1$ if $P_i = Q \circ (u, v) \circ Q'$ and $P_{1-i} = Q \circ (u, w, v) \circ Q'$ for some triangle $uvw \in X(2)$ and $i = 0, 1$.

Let $\sim$ be the smallest equivalence relation that contains the above relations (i.e. the transitive closure of two relations). We comment that these are also the relations defining the fundamental group $\pi_1(X, v_0)$ of the complex X (see e.g. [95]).

We also want to keep track of the number of (TR) equivalences used to get from one path to another. Towards this we denote by $P \sim_1 P'$ if there is a sequence of paths $(P = P_0, P_1, \dots, P_m = P')$ and $j \in [m-1]$ such that:

¹⁶If $P = (v_0, v_1, \dots, v_m)$ and $Q = (u_0 = v_m, u_1, \dots, u_n)$ then their concatenation is $P \circ Q = (v_0, v_1, \dots, v_m, u_1, \dots, u_n)$.

1. $P_j \stackrel{(TR)}{\sim} P_{j+1}$ and

2. For every $j' \neq j$, $P_{j'} \stackrel{(BT)}{\sim} P_{j'+1}$.

I.e. we can get from P to P' by a sequence of equivalences, where exactly one equivalence is by (TR) .

Similarly, for $i > 1$ we denote by $P \sim_i P'$ if we get from P to P' using i triangles.

A contraction is a sequence of equivalent *cycles* that ends with a single point. More formally,

► **Definition 75.** Let $C = (v_0, v_1, \dots, v_t, v_0)$ be a cycle around v_0 . A contraction S is a sequence of cycles $S = (C_0 = C, C_1, \dots, C_m = (v_0))$ such that $C_i \sim_1 C_{i+1}$ for every $i = 0, 1, \dots, m-1$. If the contraction contains m cycles, we say it uses m -triangles, or simply that it is an m -contraction.

A m -contraction can be viewed as a witness to $C_0 \sim_m (v_0)$.

Before giving the definition of cones which are used to prove coboundary expansion, we provide some intuition behind the connection between coboundary expansion and cones. In order to prove Equation (5.4) for a complex X , we need to upper bound the distance from any function $f \in C^1(X, \Gamma)$ to the codespace $B^1(X, \Gamma)$ (with δ_0 being the encoding function). This is done by constructing a message $g \in C^0(X, \Gamma)$ from f and $\mathcal{P}$ a family of paths from one vertex v_0 to every other vertex. Note that every edge together with the two paths in $\mathcal{P}$ from v_0 to the edge's endpoints form a cycle. The ratio between the probability that f fails the test and the distance from f to the codespace is bounded by these cycles' length of contraction. The vertex v_0 , the family of paths, and the contractions of the induced cycles constitutes a cone. The connection between the cone and coboundary expansion of X is captured by Theorem 78 and a detailed proof can be found in [29].

► **Definition 76 (cone).** A cone is a triple $\mathcal{A} = (v_0, \{P_u\}_{u \in X(0)}, \{S_{uw}\}_{uw \in X(1)})$ such that

1. $v_0 \in X(0)$.

2. For every $u \in X(0) \setminus \{v_0\}$, P_u is a walk from v_0 to u . For $u = v_0$, we take $P_{v_0} = (v_0)$ to be the cycle with no edges from v_0 to itself.

3. For every $uw \in X(1)$, S_{uw} is a contraction of $P_u \circ (u, w) \circ P_w^{-1}$.

The sequence S_{uw} depends on the direction of the edge uw . We take as a convention that S_{wu} just reverses all cycles of S_{uw} (which is also a contraction of the reverse cycle). Thus for each edge it is enough to define one of S_{uw}, S_{wu} .

The maximum length of contraction of S_{uw} among all edges uw is called the cone area.

► **Definition 77 (Cone area).** Let $\mathcal{A}$ be a cone. The area of the cone is

$$\text{Area}(\mathcal{A}) = \max_{uw \in X(1)} |S_{uw}| - 1.$$

We note that in previous works this was known as the *diameter* or *radius* of a cone [68, 29]. We chose the term “area” here since it aligns with the geometric intuition behind triangulation.

Gromov is the first to use the area of the cone to prove coboundary expansion for abelian groups. The following theorem which generalizes Gromov's approach to general groups is by [29].

► **Theorem 78** ([29, Lemma 1.6]). Let X be a simplicial complex such that $\text{Aut}(X)$ is transitive on k -faces. Suppose that there exists a cone $\mathcal{A}$ with area R . Then X is a $\frac{1}{\binom{k+1}{3} \cdot R}$ -coboundary expander for any coefficient group Γ .

5.3 Contraction via van Kampen diagrams

In this subsection we prove a version of van Kampen's lemma for our “robust” context. Previously, we defined the area of a cycle in the cone to be the length of its contraction sequence. The van Kampen lemma reduces the problem of bounding the contraction length of a cycle to counting the number of faces in a planar graph.

The van Kampen diagrams are an essential tool in geometric group theory [80, Chapter V]. They were introduced by van Kampen in 1933 in the context of the *word problem* in groups defined by presentations [97], and are a powerful visual tool for exhibiting equivalences of words of some given group presentation. They are also used to bound the number of relations required to prove that a word is equal to identity (such a bound is called the *Dehn function* of a presentation). The van Kampen diagrams are plane graphs that embed into the presentation complex of a group. The van Kampen lemma states that a word is equivalent to the identity if and only if it forms a cycle in the presentation complex which can be tiled by faces in a van Kampen diagram. Furthermore the number of inner faces of the diagram provides an upper bound on the relations needed to prove that the word equals to identity. The advantage of these diagrams is that they bound the length of a sequential process by the size of static objects.

We need a similar tool in the context of abstract complexes, which is a simple generalization of the above setup. Our main motivation for this statement is give a simple and intuitive way to bound the area of cones. Previous works that use the cones, give sequences of contractions, explaining which triangle goes after another (see e.g. [29]). This is sufficient for the proof. However, in all known examples, the ordering is just a formality, and the proofs were really describing a tiling (i.e. the contraction was really just a set of faces that formed a disk whose boundary is the cycle we needed to contract). So in this section we give justification to just drawing out the tiling via a van Kampen diagram instead of giving the full sequence of contraction.

A plane graph $H = (V, E)$ is a planar graph together with an embedding. We will denote by H both the abstract graph and the embedding.

The van Kampen lemma has a few components. First we have our simplicial complex X and a cycle C_0 in X . In addition, we have a *diagram*. This *diagram* contains:

1. A plane graph $H = (V, E)$ that is 2-vertex connected.
2. A labeling $\psi : V \rightarrow X(0)$ such that for every edge $\{u, v\} \in E$, either $\psi(u) = \psi(v)$, or $\{\psi(u), \psi(v)\} \in X(1)$.

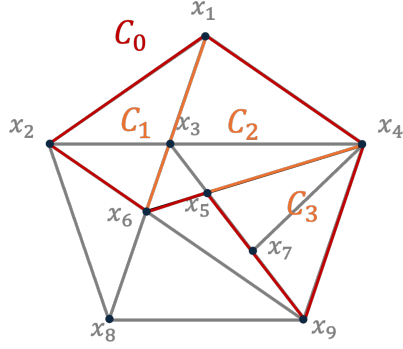
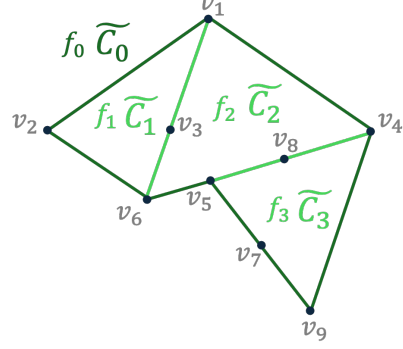
It is easy to see that a walk in H is mapped by ψ to a walk in X (possibly with repeated vertices).

Let $F = \{f_0, f_1, \dots, f_\ell\}$ be the set of faces of H with f_0 being the *outer* face. For every face let $\tilde{C}_i$ be the cycle bounding f_i , and let $C_i = \psi(\tilde{C}_i)$. To be completely formal one needs to choose a starting point and an orientation to the cycle, but we ignore this point since the choice is not important to our purposes.

We are now ready to state our van Kampen lemma.

► **Lemma 79.** *Let X, C_0, H , and F be as above. Assume that for every $i = 1, 2, \dots, \ell$ the cycle C_i has an m_i -triangle contraction. Then there exists an m_0 -triangle contraction for C_0 where $m_0 = \sum_{i=1}^{\ell} m_i$.*

A proof of the lemma can be found in Section G. Here we give a simply example to demonstrate how to use this lemma. Consider the 2-dimensional complex X in Figure 1a with 9 vertices and 10 2-faces (all triangles in the underlying graph are 2-faces in X). Let $C_0 = (x_1, x_2, x_6, x_5, x_7, x_9, x_4, x_1)$ be a cycle in X .


 (a) A 2-dimensional complex X .

 (b) A van Kampen diagram H, F of X .

 ■ **Figure 1** A van Kampen diagram for X, C_0 .

A van Kampen diagram H, F for X, C_0 is given in Figure 1b. H has 9 vertices and 4 faces $F = \{f_0, f_1, f_2, f_3\}$. We can easily check that H is 2-vertex connected and has the natural labeling $\psi(v_i) = x_i$ for $i \in [9] \setminus \{8\}$ and $\psi(v_8) = x_4$. The labeling satisfies that $\psi(\tilde{C}_i) = C_i$ for $i = 0, 1, 2, 3$.

Furthermore, it's clear from the figure that C_1, C_2 , and C_3 have 2-triangle, 3-triangle, and 2-triangle contractions respectively. So by the van Kampen lemma, C_0 has a $2 + 3 + 2 = 7$ -triangle contraction.

5.4 Cone areas and isoperimetric inequalities

In this subsection we explain the connection between cone areas and isoperimetric inequality. In a 2-dimensional simplicial complex, an isoperimetric inequality is an upper bound on the area of a cycle in terms of its length. In our context this is the following.

► **Definition 80** (Isoperimetric inequality). *Let X be a 2-dimensional simplicial complex and let $R > 0$. We say that X admits an R -isoperimetric inequality if for every cycle C with ℓ edges there exists an $R\ell$ -triangle contraction*

$$\text{Area}(C) \leq R \cdot \text{Len}(C).$$

Isoperimetric inequalities are particularly useful in presentation complexes of Cayley graphs in geometric and combinatorial group theory [80]. Cones are closely related to isoperimetric inequalities as was already observed by [52].

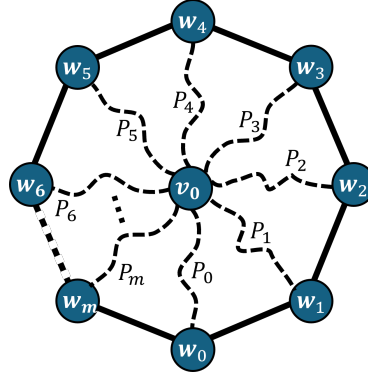
We formalize the connection with this proposition.

► **Proposition 81.** *Let X be a simplicial complex with diameter D and let $R > 0$. Then*

1. *If X admits an R -isoperimetric inequality, then X has a cone whose area is at most $(2D + 1)R$.*
2. *If X has a cone with area a then X admits an a -isoperimetric inequality.*

Here the diameter of a complex is the diameter of the underlying graph.

Sketch. For the first item, we construct a cone as follows. We fix an arbitrary vertex $v_0 \in X(0)$, and as for the paths P_u , we take arbitrary shortest paths from v_0 to u . Any cycle of the form $P_u \circ (u, w) \circ P_w^{-1}$ (where $\{u, w\} \in X(1)$) has at most $2D + 1$ edges. We take the smallest contraction of this cycle as S_{uw} . By the isoperimetric inequality, S_{uw} uses at most $(2D + 1)R$ triangles.



■ **Figure 2** The second item of Proposition 81

As for the second item, let $\mathcal{A}$ be a cone with area a and let C be a cycle with $m+1$ edges $(w_0, w_1, \dots, w_m, w_0)$. Let $P_0, P_1, \dots, P_m$ be the corresponding paths from v_0 to each of the vertices in $w_0, w_1, \dots, w_m$. By Lemma 79, we can contract C via the van Kampen diagram in Figure 2. We note that the boundary cycle of every inner face in the diagram corresponds to $P_{w_i} \circ (w_i, w_{i+1}) \circ P_{w_{i+1}}$ (where $i \in \mathbb{Z}/(m+1)$) therefore its area is at most a . We deduce the isoperimetric inequality $\text{Area}(C) \leq a \cdot (m+1) = a \cdot \text{Len}(C)$. ◀

5.5 Substitution and composition

Later in Section 6.1, we will reduce the question of bounding the cone area of some complex X to bounding that of X' which contains X as a subcomplex. The relevant definitions and auxiliary proposition for this approach are given below.

► **Definition 82.** Let X' be any complex and let X be a subcomplex of X' with $X(0) = X'(0)$. A substitution is a set $\mathcal{P} = \{P_{uv} \mid uv \in X'(1)\}$ of paths P_{uv} from u to v in X such that $P_{uv} = P_{vu}^{-1}$ (so it suffices to give one path for each pair of vertices). We say that $\mathcal{P}$ is an m -substitution if all paths have length at most m .

We note that the underlying complexes X', X of a substitution are implicit in the notation, but they will always be clear in context. Via the substitution, every path in X' has a corresponding path in X :

► **Definition 83.** Let X', X be as above and let $\mathcal{P} = \{P_{uv} \mid uv \in X'(1)\}$ be a substitution. Let $Q = (v_0, v_1, \dots, v_m)$ be a path in X' . The composition $\tilde{Q}$ is the path

$$\tilde{Q} = P_{v_0 v_1} \circ P_{v_1 v_2} \cdots \circ P_{v_{m-1} v_m} \text{ in } X.$$

The following proposition captures the key property of this substitution.

► **Proposition 84.** Let X' be any complex and let X be a subcomplex of X' containing all vertices of X' . Let $\mathcal{P}$ be a substitution. Assume that for every triangle $t = (u, v, w) \in X'(2)$, the composition $\tilde{C}_t$ of the 3-cycle $C_t = (u, v, w, u)$ has an m -triangle contraction in X . Then for every cycle C in X' that has an ℓ -triangle contraction in X' , the composition cycle $\tilde{C}$ in X has an ℓm -triangle contraction.

To prove this proposition, we need the following claim about the contraction property of two cycles with shared vertices.

▷ **Claim 85.** Let $C_s = P_1 \circ Q^{-1}$ and $C_t = P_2 \circ Q^{-1}$ be two cycles in X such that Q is nonempty. If their difference cycle $C_\Delta = P_1 \circ P_2^{-1}$ has an m -contraction, then there is an m -contraction from C_s to C_t .

We delay the proof of this claim and prove the proposition assuming this claim.

Proof of Proposition 84. Let $S = (C_0 = C, C_1, \dots, C_\ell)$ be the contraction in X' . Let $\tilde{C}_0, \tilde{C}_1, \dots, \tilde{C}_\ell$ be their compositions in X . If $C_i \sim_1 C_{i+1}$ via a TR relation $(u, v) \leftrightarrow (u, w, v)$ and some BT relations in X' , by Claim 85 $\tilde{C}_i \sim_m \tilde{C}_{i+1}$ via the m -triangle contraction of $P_{uv} \circ P_{uw} \circ P_{wv}$ in X . Since there are ℓ (TR) steps in the contraction of C , we have $\tilde{C} \sim_{\ell m} \tilde{C}_\ell$. Lastly, since C_ℓ is a point, by definition of composition so is $\tilde{C}_\ell$. So the proposition is proven. ◀

Proof of Claim 85. We first contract C_s to the intermediate cycle $P_1 \circ P_2^{-1} \circ P_2 \circ Q^{-1} = C_\Delta \circ C_t$ via (BT) relations. Then we contract C_Δ to a point without touching the part C_t . This procedure gives a m -contraction from C_s to C_t . ◁

6 Coboundary expansion

Our main result in this section is proving that the Johnson complexes $X_{\varepsilon, n}$ and their vertex links are coboundary expanders via Theorem 78. In addition, in Section 6.3, we also show that the vertex links of the complexes constructed by [50] are coboundary expanders.

► **Theorem 86.** *Let $n \in \mathbb{N}$ and $\varepsilon > 0$ satisfy that $4 \mid \varepsilon n$. Then the 2-dimensional Johnson complex X_ε is a 1-dimensional $\Omega(\varepsilon)$ -coboundary expander.*

To show the statement via Theorem 78, it boils down to show that $\text{Aut}(X_\varepsilon)$ is transitive on 2-faces and that there exists a cone $\mathcal{A}$ with area $\Theta(\frac{1}{\varepsilon})$. The first claim is straightforward. The automorphism group $\text{Aut}(X_\varepsilon)$ contains all permutations of coordinates and all translations by even-weight vectors. In particular, it is transitive on all triangles. So it is enough to bound the cone area of X_ε using the following lemma.

► **Lemma 87.** *The Johnson complex X_ε has a cone $\mathcal{A}$ with area $\Theta(\frac{1}{\varepsilon})$.*

We note that by Proposition 81, this immediately implies the following corollary.

► **Corollary 88.** *The complex X_ε admits an $\Theta(\frac{1}{\varepsilon})$ -isoperimetric inequality.*

In addition, we also show that every vertex link of X_ε is a coboundary expander.

► **Theorem 89.** *Let $n \in \mathbb{N}$ and $\frac{1}{4} \geq \varepsilon > 0$ satisfies that $8 \mid \varepsilon n$. In the 3-dimensional Johnson complex X_ε , every vertex link $(X_\varepsilon)_v$ is a 1-dimensional $\frac{1}{85}$ -coboundary expander.*

► **Remark 90.** Experts may recall the local-to-global theorem in [44] which states that if a complex is both a good local-spectral expander and its vertex links are 1-dim coboundary expanders then the complex is also a 1-dim coboundary expander. So one may suspect that the coboundary expansion theorem Theorem 86 can be derived from coboundary expansion of links (Theorem 89) and local-spectral expansion of X_ε (Lemma 59). However, this approach fails since the local-spectral expansion of X_ε is not good enough for applying the local-to-global theorem. So instead, we prove coboundary expansion of X_ε by directly using the cone method.

By link isomorphism, we can focus on the link of $\mathbf{0}$ without loss of generality. Then the automorphism group $\text{Aut}(X_{\mathbf{0}})$ contains all permutations of coordinates and so is transitive over $X_{\mathbf{0}}(2)$ (Lemma 60). By Theorem 78 it is enough to show that there exists a cone with area 85.

► **Lemma 91.** *Let $0 < \varepsilon \leq \frac{1}{4}$. Any vertex link $(X_\varepsilon)_v$ has a cone with area at most 85.*

As above, by Proposition 81, this implies:

► **Corollary 92.** *A link of a vertex $v \in X_\varepsilon$ admits an 85-isoperimetric inequality.*

6.1 Constructing a bounded area cone for X_ε

In this subsection we give a proof of Lemma 87. We take a two-step approach to showing that X_ε has a cone of bounded area. First we define a related complex X'_ε which contains X_ε as a subcomplex, and construct a cone for X'_ε with bounded area. Next, we replace every edge in X'_ε with a path in X_ε between the same two endpoints. Then effectively the edges forming a triangle in X'_ε become a cycle in X_ε . The second step is then to find a bounded-size tiling for every such cycle inside X_ε .

The key proposition used in this approach is as follows.

► **Proposition 93** (Restatement of Proposition 84). *Let X' be any complex and let X be a subcomplex of X' containing all vertices. Let $\mathcal{P}$ be a substitution. Assume that for every triangle $t = \{u, v, w\} \in X'(2)$, the composition $\tilde{C}_t$ of the 3-cycle $C_t = (u, v, w)$ has an m -triangle contraction in X . Then for every cycle C in X' that has an ℓ -triangle contraction in X' , the composition cycle $\tilde{C}$ in X has an ℓm -triangle contraction.*

This proposition implies that if for the boundary cycle C_t of every triangle $t \in X'(2)$, its composition $\tilde{C}_t$ has a $\Theta(1)$ -triangle contraction in X , then the cone area of X is asymptotically the same as that of X' . This lemma suggests a two-step strategy towards showing Lemma 87. First we construct a X'_ε which has a cone of area $\Theta(1/\varepsilon)$ (Claim 94 is proved in Section 6.1.2), and then we show that the boundary of 2-faces in X'_ε has $\Theta(1)$ -contraction in X_ε (Claim 95 is proved in Section 6.1.3).

▷ **Claim 94.** X'_ε has a cone of area $\leq 16/\varepsilon$.

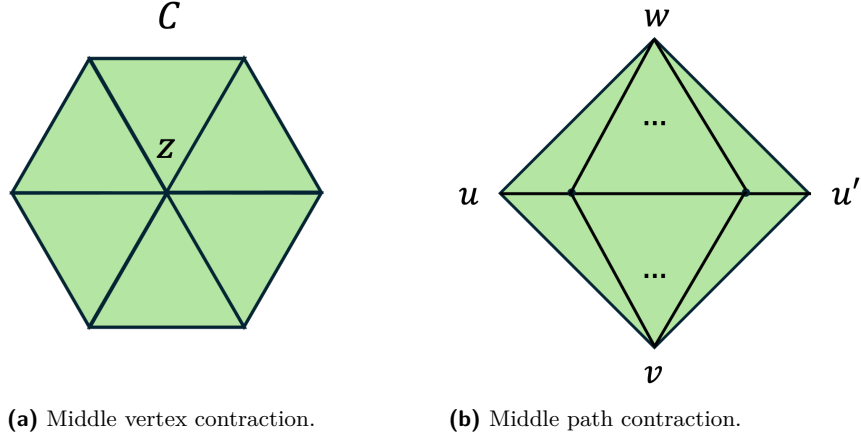
▷ **Claim 95.** For every triangle $t = \{u, v, w\} \in X'_\varepsilon(2)$ whose corresponding boundary cycle is $C_t = (u, v, w, u)$, we can contract the composition $\tilde{C}_t$ in X_ε using $O(1)$ triangles.

Postponing the proofs of these lemma and claims till later subsections, we are ready to derive Lemma 87.

Proof of Lemma 87. Claim 94 implies that the complex X'_ε has cone area $O(\frac{1}{\varepsilon})$, while Claim 95 implies that for every triangle boundary C in X'_ε , its composition $\tilde{C}$ has a $O(1)$ -triangle contraction in X_ε . By Proposition 93 we conclude that the cone area of X_ε is $O(\frac{1}{\varepsilon})$. ◀

Before moving on, we consider two “patterns” of contractions shown in Figure 3a and Figure 3b. They are going to be used extensively in the later proofs.

► **Definition 96** (Middle vertex contraction.). *A cycle C has a middle vertex contraction in a simplicial complex X if there is a vertex z so that for every edge $\{u, w\} \in C$, $\{z, u, w\} \in X(2)$.*



■ **Figure 3** Two contraction patterns.

▷ **Claim 97 (Middle vertex pattern).** Let C be a length m cycle in a simplicial complex X . If C has a middle vertex contraction, then C has an m -triangle contraction.

Proof. Given such a cycle C and a center vertex z , we can construct a van-Kampen diagram (H, ψ) for the triangles between z and edges in C , where H is the plane graph given by Figure 3a and ψ is the identity map. Note that H has m inner faces each corresponding to a triangle and an outer face whose boundary is C . Then by Lemma 79, C has an m -triangle contraction. $\triangleleft$

► **Definition 98 (Middle path contraction.).** A 4-cycle $C = (v, u, w, u')$ has a length m middle path contraction in a simplicial complex X if there is a length m walk $(u = u_1, u_2, \dots, u_m = u')$ so that for every edge $\{u_i, u_{i+1}\}$, the triangles $\{u_i, u_{i+1}, v\}, \{u_i, u_{i+1}, w\} \in X(2)$.

▷ **Claim 99 (Middle path pattern).** Let $C = (v, u, w, u')$ be a 4-cycle in a simplicial complex X . If C has a length m middle path contraction, then C has a $2m$ -contraction.

Proof. We again construct a van-Kampen diagram (H, ψ) for the set of triangles $\{u_i, u_{i+1}, v\}, \{u_i, u_{i+1}, w\} \in X(2)$, where H is the plane graph given by Figure 3b and ψ is the identity map. Note that H has $2m$ inner faces each corresponding to a triangle and an outer face whose boundary is C . Then by Lemma 79, C has an $2m$ -triangle contraction. $\triangleleft$

6.1.1 X'_ε and its cone $\mathcal{A}'$

In this part, we construct a 2-dimensional simplicial complex X'_ε such that X'_ε has a 2-substitution in X_ε .

► **Definition 100.** Let $S_{\leq \varepsilon} = \{x \in \mathbb{F}_2^n \mid wt(x) \leq \varepsilon n, wt(x) \text{ is even}\}$. The 2-dimensional complex X'_ε is given by

$$X'_\varepsilon(0) = X(0),$$

$$X'_\varepsilon(1) = \{\{x, x + s\} \mid x \in X(0), s \in S_{\leq \varepsilon}\},$$

$$X'_\varepsilon(2) = \{\{x, x + s_1, x + s_2\} \mid x \in X(0), s_1, s_2, s_1 + s_2 \in S_{\leq \varepsilon}\}.$$

▷ **Claim 101.** X'_ε has a 2-substitution in X_ε .

Proof. For any edge $(x, x+s) \in X'_\varepsilon$ (where $s \in S_{\leq \varepsilon}$), find $s_1, s_2 \in S_\varepsilon$ such that $s = s_1 + s_2$. Such s_1 and s_2 can be constructed as follows: first split $s = t_1 + t_2$ into two vectors of equal weight. Then take any vector r of weight $\varepsilon n - \frac{|s|}{2}$ with disjoint support from s , and set $s_i = r + t_i$. Now by setting $P_{x, x+s} = (x, x+s_1, x+s_1+s_2 = x+s)$, we obtain a 2-substitution for $X'_\varepsilon(1)$ in X_ε . $\triangleleft$

To bound the cone area of X'_ε , we start by constructing a cone $\mathcal{A}'$ for X'_ε . In the rest of the section we drop the subscript ε and simply write X and X' . Now we introduce a few definitions in order to describe $\mathcal{A}'$ concisely.

► **Definition 102** (Monotone path). *A path $P = (v_0, v_1, v_2, \dots, v_m)$ in X' is called monotone if $v_i \subseteq v_{i+1}$, where we identify the string with the set of coordinates that are equal to 1.*

Analogously we can define monotone cycles.

► **Definition 103** (Monotone cycle). *A cycle C is monotone if it can be decomposed to $C = P_1 \circ P_2^{-1}$ where P_1, P_2 are monotone paths of equal length.*

► **Definition 104** (Lexicographic path). *Fix an ordering on the coordinates of vectors in $X(0)$. A path $P = (v_0, v_1, v_2, \dots, v_m)$ in X' is called lexicographic if the path is monotone and every element in $v_{i+1} \setminus v_i$ is larger than all elements in v_i under the given ordering.*

We also define half-step lexicographic path.

► **Definition 105** (Half-step lexicographic path). *A path $P = (v_0, v_1, v_2, \dots, v_m)$ in X' is called a half-step lexicographic path if it is a lexicographic path from v_0 to v_m such that all but the last edge $\{v_i, v_{i+1}\}$ in the path satisfy that $|v_{i+1} \setminus v_i| = \frac{\varepsilon n}{2}$, while the last edge satisfies $|v_m \setminus v_{m-1}| \leq \frac{\varepsilon n}{2}$.*

Note that if v_0 and v_m are connected then there is a unique half-step lexicographic path between them.

Analogously we can define half-step lexicographic cycles.

► **Definition 106** (Half-step lexicographic cycle). *A cycle C in X' is half-step lexicographic if it can be decomposed to $C = P_1 \circ (x, a) \circ (a, y) \circ P_2^{-1}$ where P_1, P_2 are half-step lexicographic paths of equal length, and that $a \supseteq x \cup y$.*

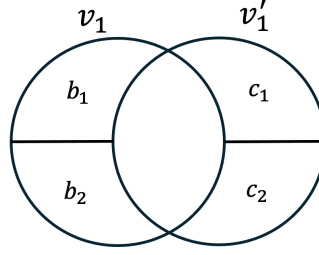
This is saying that in the half-step lexicographic cycle C , every edge $\{v_i, v_{i+1}\}$ in P_1 and P_2 satisfies that $|v_{i+1} \setminus v_i| = \frac{\varepsilon n}{2}$, while the endpoints of the two paths satisfy that $|a \setminus x| \leq \varepsilon n$ and $|a \setminus y| \leq \varepsilon n$.

We consider the following cone $\mathcal{A}'$:

► **Definition 107** (Cone of X'). *Fix an ordering on the coordinates of vectors in $X(0)$. Set the base vertex $v_0 = 0$. For every $v \in X(0) \setminus \{v_0\}$ we take P_v to be the half-step lexicographic path from v_0 to v in X' .*

6.1.2 Bounding the area of $\mathcal{A}'$: proof of Claim 94

We prove that for every $uv \in X'(1)$, the cycle $C = P_u \circ uv \circ P_v$ has a $\Theta(\frac{1}{\varepsilon})$ -triangle contraction. The strategy is to decompose C into a set of cycles in X' and compose these cycles' contractions together to obtain a contraction for C . The cycle decomposition is done by a recursive use of Lemma 79.



■ **Figure 4** Partition of $v_1 \cup v_1'$ in the proof of Claim 108.

In the rest of the section, we will use both $+$, $-$ and set operations $\setminus, \cup, \cap, ^-$ (exclusion, union, intersection, and complement) over boolean strings in $X'(0) \subset \{0, 1\}^n$. $+$, $-$ are entry-wise XOR for the boolean strings, while set operations are applied to the support of a string (i.e. the set of nonzero entries in the string).

The base case in the recursive argument is a monotone four-cycle, and we first show that all such cycles have $\Theta(1)$ -contraction.

▷ **Claim 108.** Let $C = (v_0, v_1, v_2, v_1', v_0)$ be any monotone four-cycle in X' . Then C has a ≤ 4 -triangle contraction.

Proof. By definition of monotone cycles, v_2 has size $\leq 2\epsilon n$ and that $v_1, v_1' \subseteq v_2$. If $|v_2| \leq \epsilon n$ we can contract C since $\{v_0, v_1, v_2\}, \{v_0, v_1', v_2\} \in X'(2)$.

Otherwise $\epsilon n < |v_2| \leq 2\epsilon n$, we intend to show that C has a length ≤ 2 middle path contraction. For these vertices we need to find a path between v_1, v_1' so that for every edge e in that path $e \cup \{v_0\}, e \cup \{v_2\} \in X'(2)$. Let $a = v_1 + v_1'$. We split to cases:

1. If $|a| \leq \epsilon n$ then we just take the path to be (v_1, v_1') .
2. Otherwise $\epsilon n < |a| \leq 2\epsilon n$. Then we split $v_1 \setminus v_1' = b_1 \cup b_2$ and $v_1' \setminus v_1 = c_1 \cup c_2$ both into equal-size parts as illustrated in Figure 4. So both $a_1 = b_1 + c_1$ and $a_2 = b_2 + c_2$ have size at most ϵn . We take the path $(v_1, v_1 + a_1, v_1' = v_1 + a_1 + a_2)$. Since the middle vertex $v_1 + a_1$ is contained in v_2 and has size between that of v_1 and v_1' , it is connected to v_2 and to 0.

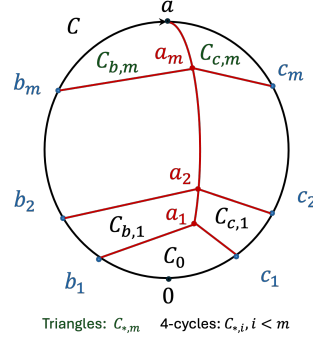
By Claim 99 there is a ≤ 4 -triangle contraction of the cycle. $\triangleleft$

The general statement for all half-step lexicographic cycles is as follows.

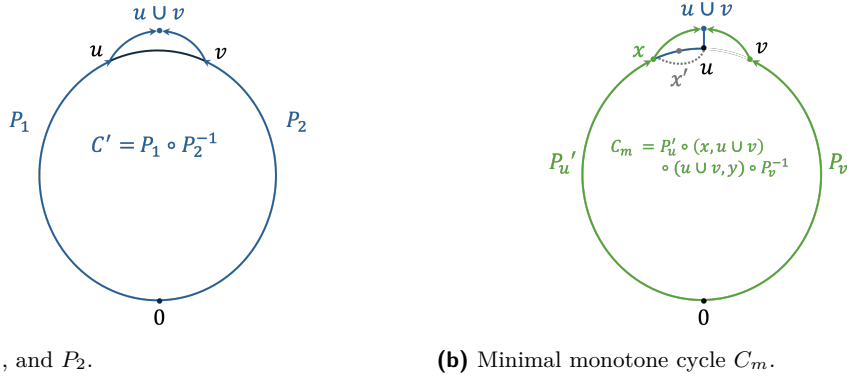
► **Lemma 109.** Let $m \geq 2$ and let $C = (0, b_1, \dots, b_m, a, c_m, \dots, c_1, 0)$ be a half-step lexicographic cycle of length $2(m+1)$. Then C has an $(8m-2)$ -contraction.

Proof. We shall decompose C into $2m-1$ monotone four-cycle and 2 triangles in X' as illustrated in Figure 5. Then applying Claim 108, we deduce that C has a $4 \cdot (2m-1) + 2 = 8m-2$ -contraction. To describe the decomposition, we define auxiliary vertices $a_i = b_i \cup c_i$ for $i = 1, \dots, m$ and $a_{m+1} = a$. We next argue that a_i is connected to b_i, c_i and a_{i+1} in X' .

First to show that for all $i \in [m]$ a_i is connected to b_i and c_i , note that $|b_i| = |c_i| = i \cdot \epsilon n / 2$. Thus if $|b_i + c_i| \leq 2\epsilon n$, then $|b_i \setminus c_i| = |c_i \setminus b_i| \leq \epsilon n$ and $a_i = b_i \cup c_i$ would be connected to both b_i and c_i . To show $|b_i + c_i| \leq 2\epsilon n$, we observe that $|c_m \setminus b_m| \leq |a \setminus b_m| \leq \epsilon n$ and analogously $|b_m \setminus c_m| \leq \epsilon n$. Now, assume for contradiction that for some $i < m$, $|b_i + c_i| > 2\epsilon n$. Similarly we deduce that $|c_i \setminus b_i|, |b_i \setminus c_i| > \epsilon n$. Suppose without loss of generality the largest element in a_i comes from b_i then $c_i \setminus b_i$ is a subset of $c_m \setminus b_m$ because any element added to b_m after the i -th step is larger than all elements in c_i . Thus $|c_m \setminus b_m| > \epsilon n$. We arrive at a contradiction. So indeed $|b_i + c_i| \leq 2\epsilon n$, and b_i and c_i are both connected to a_i for $i \in [m]$.



■ **Figure 5** Half-step lexicographic cycle decomposition in Lemma 109.



■ **Figure 6** Van-Kampen diagrams used in the proof of Claim 94.

To show that for all $i \in [m-1]$ a_i is connected to a_{i+1} , note that $a_{i+1} \supseteq a_i$ and $|a_{i+1} \setminus a_i| \leq |b_{i+1} \setminus b_i| + |c_{i+1} \setminus c_i| \leq \varepsilon n$. To see that a_m is connected to a , note that $a \supseteq a_m \supseteq b_m$ and so $|a \setminus a_m| \leq |a \setminus b_m| \leq \varepsilon n$.

Let H be the plane graph in Figure 5 whose vertices are $\{0, a_1, \dots, a_m, b_1, \dots, b_m, c_1, \dots, c_m\}$. By construction H is a van-Kampen diagram for X' with $2m+2$ inner faces.

Furthermore, we note that the cycle boundaries $C_0 = (0, b_1, a_1, c_1, 0)$, $C_{b,i} = (b_i, b_{i+1}, a_{i+1}, a_i, b_i)$ and $C_{c,i} = (c_i, c_{i+1}, a_{i+1}, a_i, c_i)$ for $i \in [m-1]$ are all monotone four cycles. $C_{b,m} = (b_m, a, a_m, b_m)$ and $C_{c,m} = (c_m, a, a_m, c_m)$ are two triangles. Applying Lemma 79 we deduce that C has a $4 \cdot (1 + 2(m-1)) + 2 = 8m - 2$ -triangle contraction. Thus we conclude the proof. $\blacktriangleleft$

Proof of Claim 94. Let $\{u, v\}$ be an edge in $X'(1)$ and $C_{uv} = P_u \circ (u, v) \circ P_v^{-1}$ be the decoding cycle. Since $\{u, v\} \in X'(1)$ it holds that $|u + v| \leq \varepsilon n$. Thereby $|u + v| = |u \cup v| - |u \cap v| \geq |u \cup v| - |v|$ and similarly $|u + v| \geq |u \cup v| - |u|$. Thus the triangle $\{v, u, u \cup v\}$ is in $X'(2)$. Define the cycle $C' = P_u \circ (u, u \cup v, v) \circ P_v^{-1}$, and the monotone paths $P_1 = P_u \circ (u, u \cup v)$ and $P_2 = P_v \circ (v, v \cup u)$ (and $C' = P_1 \circ P_2^{-1}$). If both paths have equal length, then we can contract the cycle C' using $(16/\varepsilon - 2)$ -triangles by Lemma 109 as follows. We begin by constructing the following van-Kampen diagram Figure 6a to decompose C_{uv} into C' and $C'' = (u, u \cup v, v, u)$: let H be the graph with vertices $V(P_u) \cup V(P_v) \cup \{u, v, u \cup v\}$ and edges $E(P_u) \cup E(P_v) \cup \{\{u, v\}, \{u, u \cup v\}, \{v, u \cup v\}\}$, where $V(P)$ denote P 's vertex set and $E(P)$

its edge set. By definition, the plane graph H given in Figure 6a is a van-Kampen diagram with 3 plane graph faces. H embeds into X' as illustrated by Figure 6a. In particular, the edge boundaries of the 3 faces in H are mapped to C_{uv} , C' , and C'' in X' . Therefore by Lemma 79, C_{uv} has a $(16/\varepsilon - 1)$ contraction.

If P_1 and P_2 are of different length, we do as follows (Figure 6b). Assume that P_1 is longer. Since $|u + v| \leq \varepsilon n$, the length of P_u and the length of P_v differ by at most 2. By construction, the crefix P'_u that is obtained by removing the last $|P_u| - |P_v|$ vertices and ends with a vertex x has $|x| \geq |v|$. Thus $|u \cup v \setminus x| \leq |u \cup v \setminus v| \leq \varepsilon n$. In this case, the triangle $\{x, u, v \cup u\} \in X'(2)$ can be used to decompose C_{uv} . Denote the monotone cycle $C_m = P'_u \circ (x, u \cup v) \circ (u \cup v, y) \circ P_v^{-1}$. Again by Lemma 109, C_m has a $8 \cdot (2/\varepsilon - 1) - 2$ -triangle contraction. Now we add $(x, u \cup v)$ (and also (x, u) if it is not in P_u) to the graph H . Now the van-Kampen diagram (H, Id) has 3 faces (or 4 if (x, u) is not in P_u) whose face boundaries are mapped to C_{uv} , C_m , $(x, u, u \cup v, x)$ (and potentially (x, x', u, x)). Since the last 1 (or 2) cycle has a 1-triangle contraction, C_{uv} has a $8 \cdot (2/\varepsilon - 1)$ -triangle contraction by Lemma 79. $\triangleleft$

6.1.3 Contracting $X'(2)$ with X : proof of Claim 95

We move back to consider the Johnson complex X which is a subcomplex of X' .

▷ **Claim 110.** Let $C = (v_0, v_1, v_2, v'_1, v_0)$ be *any* four-cycle in X . Then C has an 8-triangle contraction.

Note that in contrast to Claim 108 we do not assume monotonicity.

Proof. Without loss of generality $v_0 = 0$. In this case v_2 has $|v_2| = t$ for some even $2 \leq t \leq 2\varepsilon n$. We intend to find a path between v_1 and v'_1 and then apply the middle path contraction from Claim 99. For this we define the following graph $G = (V, E)$.

$$V = N(0) \cap N(v_2)$$

$$E = \{\{u, w\} \mid \{u, w, 0\}, \{u, w, v_2\} \in X(2)\}.$$

We note that $v_1, v'_1 \in V$, and that if there is a length- ℓ path from v_1 to v'_1 in this graph, then the four-cycle has a 2ℓ -triangle contraction by Claim 99.

To find such paths, we define two functions mapping tensor products of Johnson graphs to G .

$$\psi_c : G_c := J(t, \frac{t}{2}, \lceil \frac{t}{4} \rceil) \otimes J(n - t, \varepsilon n - \frac{t}{2}, \frac{\varepsilon}{2}n - \lceil \frac{t}{4} \rceil) \rightarrow G,$$

$$\psi_f : G_f := J(t, \frac{t}{2}, \lfloor \frac{t}{4} \rfloor) \otimes J(n - t, \varepsilon n - \frac{t}{2}, \frac{\varepsilon}{2}n - \lfloor \frac{t}{4} \rfloor) \rightarrow G,$$

where every vertex (u, w) in the two tensor graphs where $u \in \binom{v_2}{t/2}$ and $w \in \binom{[n] \setminus v_2}{\varepsilon n - t/2}$ is mapped to $\psi_c((u, w)) = \psi_f((u, w)) = u \cup w$. Here we implicitly identify the t elements in $J(t, \frac{t}{2}, *)$ with the set v_2 , and the $n - t$ elements in $J(n - t, \varepsilon n - \frac{t}{2}, *)$ with $[n] \setminus v_2$. Note that the two tensor product graphs have the same vertex space. Furthermore when $\frac{t}{4}$ is an integer, the two functions are identical.

We first show that the functions give two bijections between the vertex sets: for any distinct vertices $(u, w), (u', w')$ in the tensor product graphs, the unions $u \cup w$ and $u' \cup w'$ are also distinct in V . Furthermore, for any $v \in V$, v and $v_2 - v$ both have weight εn . This implies that $|v \cap v_2| = \frac{t}{2}$ and $|v \setminus v_2| = \varepsilon n - \frac{t}{2}$. Therefore v has the preimage $(v \cap v_2, v \setminus v_2)$ under ψ_c . Vertex bijectivity follows.

We next show that ψ_c is actually a graph homomorphism. That is to say every edge in G_c is mapped to an edge in G . For every edge $\{(u_1, w_1), (u_2, w_2)\}$ in G_c , we have $|u_1 \cap u_2| = \lceil \frac{t}{4} \rceil$ and $|w_1 \cap w_2| = \frac{\varepsilon n}{2} - \lceil \frac{t}{4} \rceil$. Subsequently

$$\psi_c((u_1, w_1)) + \psi_c((u_2, w_2)) = (u_1 \cup w_1) + (u_2 \cup w_2) = (u_1 + u_2) \cup (w_1 + w_2)$$

has weight $2(\frac{t}{2} - \lceil \frac{t}{4} \rceil) + 2(\varepsilon n - \frac{t}{2} - (\frac{\varepsilon n}{2} - \lceil \frac{t}{4} \rceil))\varepsilon n = \varepsilon n$. Therefore $\{\psi_c((u_1, w_1)), \psi_c((u_2, w_2))\}$ is an edge in G . Thus ψ_c is a vertex bijective graph homomorphism. A similar argument shows that the same is true for ψ_f .

We next use the following claim to find length-4 paths between any pair of vertices in $G_c \cup G_f$ and then transfer this property to G via ψ_c, ψ_f .

▷ **Claim 111.** Let $G_c = (V_c, E_c)$ and $G_f = (V_f = V_c, E_f)$ be defined as above. For any two vertices $v_1 = (u, w)$ and $\beta = (u', w')$ in V such that $u \neq u'$ and $w \neq w'$, there is a length two path $(v_1, (x, y), \beta)$ in $G_{cf} = (V_c, E_c \cup E_f)$ such that $\{v_1, (x, y)\} \in E_c$ and $\{(x, y), \beta\} \in E_f$.

The claim immediately implies that for any pair of vertices v_1, v'_1 in the union graph G_{cf} , we can find an intermediate vertex β such that the two components in β are distinct from the components in v_1 and v'_1 , and then find a length-4 path of the form $(v_1, \alpha, \beta, \alpha', v'_1)$ in G_{cf} .

We postpone the proof of the claim and conclude that using the vertex bijective graph homomorphisms ψ_c and ψ_f , such paths also exist in G . Therefore there exists a length-4 path between any two vertices v_1, v'_1 in G . So the four-cycle has an 8-triangle contraction. ◁

Proof of Claim 111. When $\frac{t}{4}$ is an integer, $G_c = G_f$. The existence of the intermediate vertex (x, y) is derived from the fact that there is a length two path between any two vertices in the Johnson graphs $J(t, \frac{t}{2}, \frac{t}{4})$ and $J(n - t, \varepsilon n - \frac{t}{2}, \frac{\varepsilon n}{2} - \frac{t}{4})$.

Now we focus on the case when $\frac{t}{2}$ is odd. Construct x as illustrated in Figure 7a. Take $a_x \subseteq u \setminus u'$ s.t. $|a_x| = \lceil \frac{|u \setminus u'|}{2} \rceil$, $b_x \subseteq u \cap u'$ s.t. $|b_x| = \lceil \frac{|u \cap u'|}{2} \rceil$, $c_x \subseteq u' \setminus u$ s.t. $|c_x| = \lceil \frac{|u' \setminus u|}{2} \rceil - 1$, and $d_x \subseteq \overline{u \cup u'}$ s.t. $|d_x| = \lceil \frac{|u \cup u'|}{2} \rceil$ ¹⁷. Then set $x = a_x \cup b_x \cup c_x \cup d_x$. We can verify that $|u \cap x| = \lceil \frac{t}{4} \rceil$, $|x \cap u'| = \lceil \frac{t}{4} \rceil - 1 = \lfloor \frac{t}{4} \rfloor$, and $|x| = \lceil \frac{t}{4} \rceil + \lfloor \frac{t}{4} \rfloor = \frac{t}{2}$.

Similarly, construct y as illustrated in Figure 7b. Take $a_y \subseteq w \setminus w'$ s.t. $|a_y| = \lfloor \frac{|w \setminus w'|}{2} \rfloor$, $b_y \subseteq w \cap w'$ s.t. $|b_y| = \lfloor \frac{|w \cap w'|}{2} \rfloor$, $c_y \subseteq w' \setminus w$ s.t. $|c_y| = \lfloor \frac{|w' \setminus w|}{2} \rfloor + 1$, and $d_y \subseteq \overline{w \cup w'}$ s.t. $|d_y| = \lfloor \frac{|w \cup w'|}{2} \rfloor$. Then set $y = a_y \cup b_y \cup c_y \cup d_y$. We can verify that $|w \cap y| = \frac{\varepsilon n}{2} - \lceil \frac{t}{4} \rceil$, $|y \cap w'| = \frac{\varepsilon n}{2} - \lceil \frac{t}{4} \rceil + 1 = \frac{\varepsilon n}{2} - \lfloor \frac{t}{4} \rfloor$, and $|y| = \lfloor \frac{|w|}{2} \rfloor + \lceil \frac{|w'|}{2} \rceil = \varepsilon n - \frac{t}{2}$.

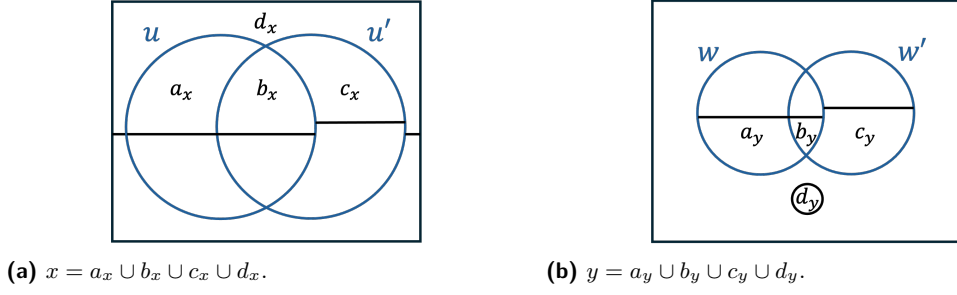
Now by construction $\{(u, w), (x, y)\} \in E_c$ and $\{(x, y), (u', w')\} \in E_f$, and thus we conclude the proof. ◁

We are ready to prove the main result of this subsection

Proof of Claim 95. Recall that by Claim 101 for every edge in $\{x, x + s\} \in X'(1)$ we can find some $s_1, s_2 \in S_\varepsilon$ such that path $P_{x, x+s} = (x, x + s_1, x + s_1 + s_2 = x + s)$. So we focus on six-cycles in X of the form $C = (u, a, v, b, w, c, u)$ such that (u, v, w) is a triangle in X' (i.e. $v - u, w - u, v - w$ all have weight at most εn). Let us further assume without loss of generality that $u = 0$. Thus, $C = (0, a, v, b, w, c, 0)$ where $v, w, v - w$ all have weight at most εn .

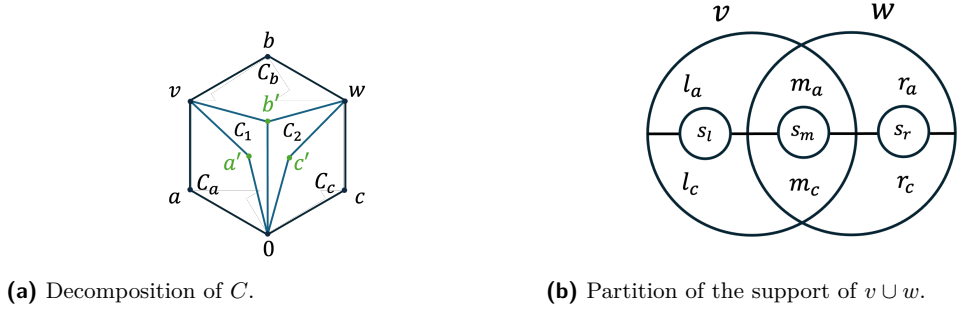
Our plan is as follows:

¹⁷Since $u \neq u'$, the sizes on the right-hand side of the size constraints are nonnegative. So these size constraints are satisfiable.



■ **Figure 7** Construction of the intermediate vertex (x, y) .

1. We find vertices a', b', c' such that $(0, a', v, b', w, c', 0)$ is a cycle in X , and such that a', b', c' are neighbors of 0.
2. We decompose C to five four-cycles C_a, C_b, C_c, C_1, C_2 as illustrated in Figure 8a. Then by Claim 110 each of the four-cycle has a 8-triangle contraction. Furthermore, the figure gives a van-Kampen diagram with C being the outer face boundary, and the four-cycles being the inner face boundaries. So applying Lemma 79, we can conclude that C has a 40-triangle contraction.



■ **Figure 8** Decomposition and partition in the proof of Claim 95.

Now we construct a', b', c' . The construction is easy when $v \cap w$ has even weight. In this case $v \setminus w$ and $w \setminus v$ also have even weight. We construct the auxiliary vertices by first evenly partitioning $v \setminus w = l_a \cup l_c$, $v \cap w = m_a \cup m_c$ and $w \setminus v = r_a \cup r_c$, then finding a set $x \subseteq \overline{v \cup w}$ s.t. $|x| = \varepsilon n - |v \cup w|/2$, and setting $a' = l_a + m_a + r_a + x$, $b' = l_a + m_a + r_c + x$, and $c' = l_c + m_c + r_c + x$. By construction the three vertices are adjacent to 0 and form the cycle $(0, a', v, b', w, c', 0)$ in X .

The construction becomes complicated when $v \cap w$, $v \setminus w$, and $w \setminus v$ have odd weights. In this case, we partially partition the supports of $v \cap w$, $v \setminus w$, and $w \setminus v$ as follows (Figure 8b):

$$\begin{aligned}
 v \setminus w &= l_a \cup l_c \cup s_l & \text{s.t. } \begin{cases} |l_a| = \lfloor |v \setminus w|/2 \rfloor \\ |l_c| = \lfloor |v \setminus w|/2 \rfloor \end{cases} \\
 v \cap w &= m_a \cup m_c \cup s_m & \text{s.t. } \begin{cases} |m_a| = \lfloor |v \cap w|/2 \rfloor \\ |m_c| = \lfloor |v \cap w|/2 \rfloor \end{cases} \\
 w \setminus v &= r_a \cup r_c \cup s_r & \text{s.t. } \begin{cases} |r_a| = \lfloor |w \setminus v|/2 \rfloor \\ |r_c| = \lfloor |w \setminus v|/2 \rfloor \end{cases}
 \end{aligned}$$

where $|s_l| = |s_m| = |s_r| = 1$.

Then we pick $a' = l_a + m_a + r_a + s_m + x$ and $c' = l_c + m_c + r_c + s_m + x$ where x is supported over $v \cup w$ and $|x| = \varepsilon n - \lfloor |v \cup w|/2 \rfloor$. Note that since $|v \cup w| = |v| + |w \setminus v| \leq |v| + |w - v| \leq 2\varepsilon n$, the cardinality $\varepsilon n - \lfloor |v \cup w|/2 \rfloor \geq 0$ is well defined. Define $b' = l_a + m_a + r_c + s_m + x$. We note that $|v - a'| = |l_c + m_c + r_a + s_l + x| = \varepsilon n$. Similarly $|w - c'| = |l_c + m_a + r_a + s_r + x| = \varepsilon n$, so a' and c' have edges to v and w respectively. Furthermore $|v - b'| = |l_c + m_c + r_c + s_l + x| = \varepsilon n$ and $|w - b'| = |l_a + m_c + r_a + s_r + x| = \varepsilon n$. So b' is connected to v and w . By construction a', b', c' are all connected to 0 therefore the cycles C_a, C_b, C_c, C_1, C_2 all exist. Thus we complete the proof. $\triangleleft$

6.2 Cone diameter of the vertex links: proof of Lemma 91

Recall that the links of every vertex in the complex are isomorphic to one another. Therefore it is enough to prove the lemma for the link of 0. By Lemma 60, the underlying graph of the link complex is $J(n, \varepsilon n, \frac{\varepsilon}{2}n)$, and the 2-faces are

$$X_0(2) = \left\{ \{v_0, v_1, v_2\} \mid \{v_0, v_1\}, \{v_1, v_2\}, \{v_0, v_2\} \in X_0(1) \text{ and } |v_0 \cap v_1 \cap v_2| = \frac{\varepsilon}{4}n \right\}.$$

Let us begin with the following claim.

$\triangleright$ **Claim 112.** Let $C = (v, u, w, u', v)$ be a 4-cycle in $J(n, \varepsilon n, \frac{\varepsilon}{2}n)$ such that $v \cap w = \emptyset$, then C has a contraction using 4 triangles.

Proof. We intend to use the middle path pattern from Claim 99. So let us show that there exists a path (u, z, u') such that both edges in the path form triangles with v and with w . To find z , we find separately $z_1 = z \cap v$ and $z_2 = z \cap w$ and set $z = z_1 \cup z_2$. To find z_1 we set $x_1 = u \cap v, x_2 = u' \cap v$ and note that both of these have size $\frac{\varepsilon}{2}n$ while v has size εn . We note that $|x_1 \setminus x_2| = |x_2 \setminus x_1|$ and we take z_1 by taking $t = \min\{|x_1 \setminus x_2|, \frac{\varepsilon}{4}\}$ coordinates from each of $x_1 \setminus x_2, x_2 \setminus x_1$, an additional $\frac{\varepsilon}{4}n - t$ coordinates from each of $x_1 \cap x_2, v \setminus (x_1 \cup x_2)$. This is possible since:

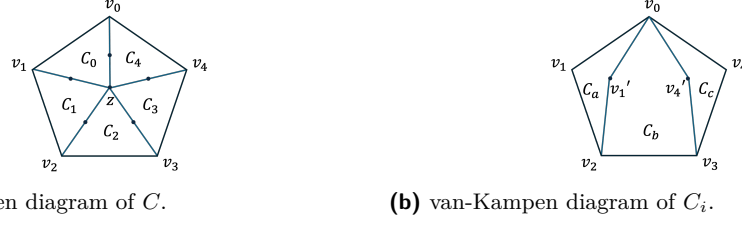
1. $|x_1 \cap x_2| = \frac{\varepsilon}{2}n - t$ and
 2. $|v \setminus (x_1 \cup x_2)| \geq \varepsilon n - |x_1 \cup x_2| = \varepsilon n - |x_2| - |x_1 \setminus x_2| \geq \frac{\varepsilon}{2}n - t$.
- Find z_2 analogously in w . Then by construction $\{u, v, z\}, \{u', v, z\}, \{u, w, z\}, \{u', w, z\}$ are all in $X_0(2)$. So Claim 99 implies that C has a 4-triangle contraction. $\triangleleft$

Proof of Lemma 91. Without loss of generality consider the link of 0. Let $v_0 \in X_0$ be the base vertex. Since the diameter of the Johnson graph is 2, for any $u \in X_0(0) \setminus \{v_0\}$ we arbitrarily choose a path of length 2 from v_0 to u as P_u (the existence of triangles ensures that we can do this for neighboring vertices as well). To bound the cone diameter, we consider any edge $\{u, u'\} \in X_0(1)$ and its corresponding cone cycle $C = P_u \circ (u, u') \circ P_{u'}^{-1}$. Let us re-annotate $C = (v_0, v_1, v_2, v_3, v_4, v_0)$.

Let us begin by reducing to the case where v_0 is disjoint from $v_2 \cup v_3$. Observe that $|v_i| = \varepsilon n$ and that $|v_i \setminus \bigcup_{j=0}^{i-1} v_j| \leq |v_i \setminus v_{i-1}| \leq \frac{\varepsilon}{2}n$. Thus $|\bigcup_{j=0}^4 v_j| \leq 3\varepsilon n$ and by assumption that $\varepsilon \leq \frac{1}{4}$ there exists a vertex z such that z is disjoint from $\bigcup_{j=0}^4 v_j$.

Since the diameter of the Johnson graph is 2, we can decompose C into 5 five-cycles C_0, C_1, C_2, C_3, C_4 via the van-Kampen diagram given in Figure 9a. Then by Lemma 79, if every C_i can be contracted by m triangles, the original cycle C has a $5m$ -triangle contraction. So we will show next that every C_i has a contraction of 17 triangles.

Re-annotate $C_i = (v_0, v_1, v_2, v_3, v_4, v_0)$ such that v_0 is disjoint from $v_2 \cup v_3$. We can find some v'_1 and v'_4 such that the following holds:



■ **Figure 9** Van-Kampen diagrams used in the proof of Lemma 91.

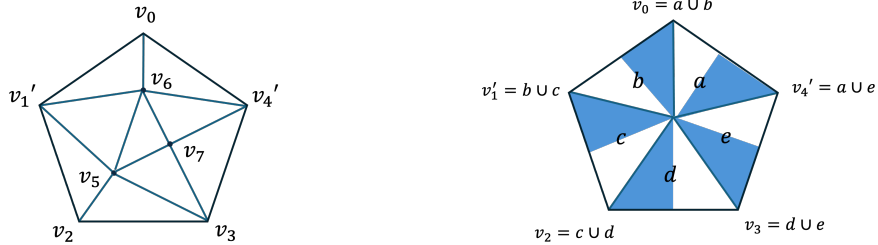
1. v'_1 is disjoint from $v_3 \cup v'_4$.
2. v'_4 is disjoint from $v'_1 \cup v_2$.
3. The path $P' = (v_0, v'_1, v_2, v_3, v'_4, v_0)$ is a closed walk in X_0 .

If such vertices exists, then C_i can be decomposed into two four-cycles C_a, C_c and a five-cycle C_b via the van-Kampen diagram in Figure 9b. Furthermore since $v_0 \cap v_2 = v_0 \cap v_3 = \emptyset$, each of the four-cycles can be contracted by 4 triangles (Claim 112). Furthermore, we will soon show that C_b has a 9-triangle contraction using the following claim.

▷ **Claim 113.** If a five-cycle $C = (v_0, v'_1, v_2, v_3, v'_4, v_0)$ in X_0 satisfies that every vertex $v \in C$ is disjoint from the cycle vertices that are not neighbors of v in C , then C has a 9-triangle contraction in X_0 .

Before proving this claim let us complete the proof of the lemma. Applying Lemma 79 again, we can conclude that C_i has a $9 + 2 \cdot 4 = 17$ -triangle contraction.

We now find such v'_1 and v'_4 . Partition v_0 into $s_1 \cup s_2$ of equal size $\frac{\varepsilon}{2}n$. Then define $v'_1 = s_1 \cup (v_2 \setminus v_3)$ and $v'_4 = s_2 \cup (v_3 \setminus v_2)$. The two sets both have size εn since $|s_1| = |s_2| = |v_2 \setminus v_3| = |v_3 \setminus v_2| = \frac{\varepsilon}{2}n$. It is straightforward to verify that the three properties above are satisfied. We thus complete the proof. ◀



■ **Figure 10** Decomposition of C in Claim 113.

Proof of Claim 113. Define sets $a, b, c, d, e \subseteq [n]$ to be $a = v'_4 \cap v_0, b = v_0 \cap v'_1, c = v'_1 \cap v_2, d = v_2 \cap v_3$ and $e = v_3 \cap v'_4$. By the assumption on the vertices, the 5 sets all have size $\frac{\varepsilon n}{2}$ and satisfy that $v_0 = a \cup b, v'_1 = b \cup c, v_2 = c \cup d, v_3 = d \cup e$, and $v'_4 = e \cup a$. Furthermore, for each $x \in \{a, b, c, d, e\}$ we partition it into two disjoint sets x_1, x_2 of size $\frac{\varepsilon n}{4}$ as illustrated in Figure 10b.

Now we can decompose C into 9 triangles as illustrated in Figure 10a by defining v_5, v_6 , and v_7 as follows: $v_5 = b_1 \cup c_1 \cup d_1 \cup e_1, v_6 = a_1 \cup b_1 \cup c_2 \cup e_1, v_7 = a_1 \cup b_1 \cup d_1 \cup e_2$. We can verify that each of the 9 inner face in the figure has its boundary being a triangle in $X_0(2)$. Thus C has a 9-triangle contraction. ◀

6.3 Coboundary expansion of links in the complexes of [50]

Even though most of this Section 6 is devoted to the Johnson complex, our techniques also extend to the complexes constructed by [50] that were described in Section 3. In this section we prove that the link of every vertex in these complexes are coboundary expanders.

► **Lemma 114.** *Let $Y = \mathcal{SP}(\mathcal{G}(\mathbb{F}_2^n)^{\leq d})$ for $n > 2d$. Let $X = \beta(Y)$ be the basification of Y . Then X has a cone with area 459 and is therefore a 1 -dimensional $\frac{1}{459}$ -coboundary expander.*

The proof of this lemma is just a more complicated version of the proof of Lemma 91. Before proving the lemma, we give an explicit description of the link structure. The vertices are $X(0) = \mathcal{M}(\frac{d}{2})$, i.e. all matrices of a certain rank $\frac{d}{2}$. The edges are

$$X(1) = \left\{ \{A_1, A_2\} \mid \exists B \in \mathcal{M}(\frac{d}{4}), B \oplus (A_1 - B) \oplus (A_2 - B) \right\}.$$

Finally, the triangles $X(2)$ are all $\{A_1, A_2, A_3\}$ such that there exist $B_1, B_2, \dots, B_7 \in \mathcal{M}(\frac{d}{8})$ such that the sum $\bigoplus_{j=1}^7 B_j$ is direct, and

$$A_1 = B_1 \oplus B_2 \oplus B_3 \oplus B_4,$$

$$A_2 = B_1 \oplus B_2 \oplus B_5 \oplus B_6,$$

$$A_3 = B_1 \oplus B_3 \oplus B_5 \oplus B_7.$$

This description is a direct consequence of the definition of admissible functions in Section 3.

We follow the footsteps of the proof of Lemma 91. Towards this we need two claims. The first proves that there exist length 4 paths between any two vertices in X , so that we can construct a cone where the cycles have constant length. This was immediate in the Johnson case but needs to be argued here. This claim also constructs length-2 paths between any two matrices that form a direct sum. we will need this later on.

► **Claim 115.**

1. For any two matrices $A, B \in X(0)$ there is a path of length 4 between A and B .
2. If $A \oplus B$ then there is a path (A, C, B) . Moreover, there exists such a path where the middle vertex $C = A_1 \oplus B_1$ such that the components of this sum have equal rank and satisfy $A_1 \leq A$ and $B_1 \leq B$ ¹⁸.

The second claim is a variant of Claim 112. Claim 112 bounds 4-cycles in the Johnson complex provided that two non-neighbor vertices in the cycle are disjoint. This claim is the analog for X :

► **Claim 116.** Let (Z, C, A, D, Z) be a 4-cycle with the following properties.

1. The matrices $Z \oplus A$ are a direct sum.
2. We can decompose $C = C_1 \oplus C_2$ and $D = D_1 \oplus D_2$ such that $C_1, D_1 \leq Z$ and $C_2, D_2 \leq A$.¹⁹

Then the cycle has a tiling with $O(1)$ -triangles.

¹⁸ Not all paths from A to B have this property. For instance, let e_1, e_2, e_3, e_4 be four independent vectors. Let $A = e_1 \otimes e_1 + e_2 \otimes e_2$, $C = e_2 \otimes e_2 + e_3 \otimes e_3$ and $B = (e_2 + e_3) \otimes e_3 + e_4 \otimes e_4$. One can verify that because $C = (e_2 + e_3) \otimes e_3 + e_2 \otimes (e_2 + e_3)$ then (A, C, D) is indeed a path. However, we can directly check that there are no $A_1 \leq A, B_1 \leq B$ such that $C = A_1 \oplus B_1$.

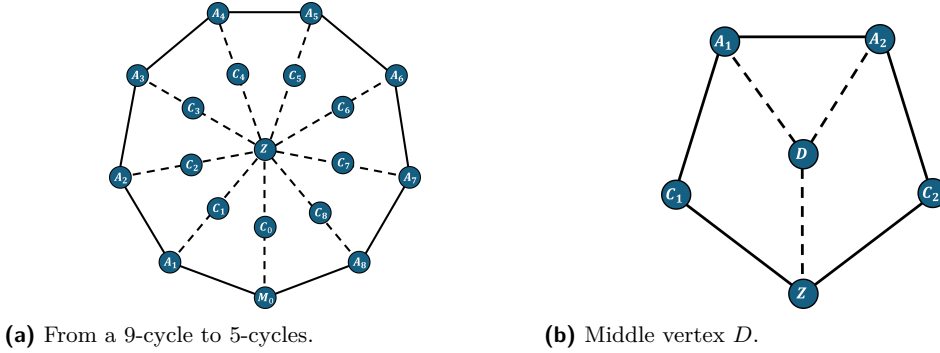
¹⁹ We note that there is no analogue to second property of the cycle in the statement Claim 112. The analogue is that $C = (C_1 \cap A) \cup (C \cap Z)$ (and similarly for D). This follows from the first property in the Johnson case but we require it explicitly in this case.

Both of these claims are proven after the lemma itself.

Proof of Lemma 114. It is easy to see that the complex X has a transitive action on the triangles. Therefore by Theorem 78, it suffices to find a cone whose area is 459 to prove coboundary expansion.

Let M_0 be an arbitrary matrix we take as the root of our cone. We take arbitrary length 4 paths P_A from M_0 to every matrix $A \in X(0)$. This is possible by Claim 115.

For every edge $\{A, A'\}$, we need to find a contraction for the cycle $P_A \circ (A, A') \circ P_{A'}^{-1}$. We note that this cycle has length 9. Therefore it suffices to show that every 9 cycle $(M_0, A_1, A_2, \dots, A_8, M_0)$ has a tiling with 459 triangles. We begin by reducing the problem into tilings of 5 cycles instead. Indeed, observe that for every edge $\{A_i, A_{i+1}\}$ in the cycle, $\dim(\text{row}(A_i) + \text{row}(A_{i+1})) = \dim(\text{col}(A_i) + \text{col}(A_{i+1})) = \frac{3d}{4}$. Therefore there exists some $Z \in X(0)$ such that $\text{row}(Z)$ and $\text{col}(Z)$ are a direct sum with $\text{row}(A_i) + \text{row}(A_{i+1})$ and $\text{col}(A_i) + \text{col}(A_{i+1})$ respectively (In fact, a random matrix Z will satisfy this with high probability for all edges in the cycle because $n \geq 2d$).



■ **Figure 11** Decomposition of the cycles.

By Claim 115 there exists a length two path from Z to every vertex A_i in the cycle as in Figure 11a. Therefore, by drawing these paths we can decompose our cycle into 9 cycles of length 5 of the form $(Z, C_i, A_i, A_{i+1}, C_{i+1}, Z)$ such that the row/column spaces of Z are a direct sum with $\text{row}(A_1) + \text{row}(A_2)$ and $\text{col}(A_1) + \text{col}(A_2)$. This is a similar step as the one we do in Figure 9a in the Johnson case.

Fix a cycle as above let us denote by B the matrix such that $B \oplus (A_1 - B) \oplus (A_2 - B)$. Let Z_1, Z_2 be of equal rank such that $Z = Z_1 \oplus Z_2$. Denote by $D = Z_1 \oplus B$. We observe that $\{Z, D\}, \{A_1, D\}$ and $\{A_2, D\}$ are all edges and therefore we can decompose this cycle into a 3-cycle (D, A_1, A_2, D) and two 4-cycles (Z, C_i, A_i, D, Z) that we need to tile (see Figure 11b).

The tiling of (D, A_1, A_2, D) is simple. We decompose $Z_1, B, A_1 - B$ and $A_2 - B$ into a direct sum of matrices of equal rank $Z_1 = K_1 \oplus K_2, B = L_1 \oplus L_2, A_1 - B = M_1 \oplus M_2$ and $A_2 - B = N_1 \oplus N_2$. Then the matrix $E = K_1 \oplus L_1 \oplus M_1 \oplus N_1$ is connected in a triangle with all three edges. so we can triangulate this three cycle using three triangles.

As for the cycles (Z, C_i, A_i, D, Z) , we tile them with the claim Claim 116 using 24-triangles. Thus we can triangulate every 5-cycle using of 51 triangles. As every 9-cycle is decomposed into 9 such 5-cycles we get a total of 459 triangles, so the cone area is 459. ◀

Proof of Claim 115. First note that the first statement follows from the second, because for any A, B there exists some $D \in X(0)$ such that $A \oplus D$ and $B \oplus D$ (this is because of the assumption that $n > 2d$). Thus we can take a 2-path from A to D and then another 2-path from D to B .

Let us thus show that assuming that $A \oplus B$ there is a 2-path from A to B . Decompose $A = A_1 \oplus A_2$ of equal rank, and similarly for $B = B_1 \oplus B_2$. Let $C = A_1 \oplus B_1$ and take the path (A, C, B) . The claim follows. $\triangleleft$

Up until now, this was largely similar to Lemma 91. Surprisingly, the proof of Claim 116 is more complicated than its Johnson analogue Claim 112. This stems from the fact the matrix domination poset is different from the matrix Johnson Grassmann poset, and in particular there isn't a perfect analogue to set intersection in the matrix poset (this is called a meet in poset theory).

Proof of Claim 116. We will use the middle path pattern Claim 99. We observe that $C_1, D_1 \leq Z$ and $D_2, D_2 \leq A$, and we can view them as vertices in $\mathcal{T}^Z$ and $\mathcal{T}^A$ respectively. We will use a similar strategy as in Claim 112, and use the fact that the graph $\mathcal{T}^Z$ has short paths between every two matrices. That is,

$\triangleright$ Claim 117. For every m and two matrices $C, D \in \mathcal{T}^m$ there is a path of length 12 between C and D .

This claim is proven in the end of Section D.4.

Using this claim, we find a pair of 12-paths from C_i to D_i in $\mathcal{T}^Z$ and $\mathcal{T}^A$ respectively: $(C_1, M^1, M^2, \dots, M^{11}, D_1)$ and $(C_2, N^1, N^2, \dots, N^{11}, D_2)$. Then we observe that the matrices $K^i = M^i \oplus N^i$ form a path $(C, K^1, K^2, \dots, K^{11}, D)$ from C to D in X . Moreover, every edge in this path is in a triangle with A and with Z . The claim is proven. $\triangleleft$

6.3.1 Coboundary expansion in the links and cosystolic expansion

In [50] it is proven that the Cayley complex whose link is X is not a coboundary expander for any constant $\beta > 0$. It is done by proving that $Z^1(X, \Gamma) \neq B^1(X, \Gamma)$.

However, we can still prove that these complexes are *cosystolic expanders*. This notion is a weakening of coboundary expansion that allows the existence of cohomology, i.e. for $Z^1(X, \Gamma) \neq B^1(X, \Gamma)$:

► Definition 118 (1-dimensional cosystolic expander). *Let X be a d -dimensional simplicial complex for $d \geq 2$. Let $\beta > 0$. We say that X is a 1-dimensional β -cosystolic expander if for every group Γ , and every $f \in C^1(X, \Gamma)$ there exists some $h \in Z^1(X, \Gamma)$ such that*

$$\beta \operatorname{dist}(f, h) \leq \operatorname{wt}(\delta f). \quad (6.1)$$

The difference between this definition and the definition of coboundary expansion, is that in the definition of coboundary expansion we require $h = \delta g$, or equivalently $h \in B^1(X, \Gamma)$ whereas here we only require $h \in Z^1(X, \Gamma)$ which is less restrictive.

We note that cosystolic expansion is still an important property and can sometimes be used in applications such as topological overlapping property [43] or to get some form of agreement testing guarantee [27].

Kaufman, Kazhdan and Lubotzky were the first to prove that in local spectral expanders, cosystolic expansion follows from coboundary expansion of the links [60] (together with local spectral expansion). Later on their seminal result was extended to all groups and higher dimensions [44, 63, 64, 65, 28]. In particular, we can use the quantitatively stronger theorem in [28, Theorem 1.2], to get cosystolic expansion from Lemma 114 and the local spectral expansion.

► **Corollary 119.** *Let $Y = \mathcal{SP}(\mathcal{G}(\mathbb{F}_2^n)^{\leq d})$ for $n > 2d$. Let $X = \beta(Y)$ be the basification of Y . Let $\text{Cay}(\mathbb{F}_2^n, S)$ be the complexes whose links are isomorphic to X , as constructed in Section 3. Assume that n is sufficiently large such that $\text{Cay}(\mathbb{F}_2^n, S)$ is a 10^{-4} -local spectral expander. Then $\text{Cay}(\mathbb{F}_2^n, S)$ is also a 1-dimensional 10^{-4} -cosystolic expander.*

We omit the details.

7 Induced Grassmann posets

In this subsection we will present a generalization of our Johnson and Matrix Grassmann posets. After a few preliminaries, we will present a way to construct Grassmann posets based on the Hadamard encoding. We will see that such complexes have a simple description of their links. Afterwards, we will show that on some of these complexes we can decompose the links into a tensor product of simple graphs, generalizing some of the previous results in this paper. We start with some classical facts about quotient spaces and bilinear forms.

7.1 Preliminaries on quotient spaces and bilinear forms

Quotient spaces

Let $U \subseteq V$ be a subspace. The quotient space is a vector space V/U whose vectors are the sets $v + U = \{v + y \mid y \in U\}$ for every vector $v \in V$. It is standard that addition (and scalar multiplication) is well defined on these sets. Addition is defined as $(v + U) + (x + U) := (v + x) + U$ and the sum of the two sets does not depend on the representatives $v \in v + U$ and $x \in x + U$.

Bilinear forms

Let V_1, V_2 be two vector spaces. A bilinear form is a function $\langle \rangle : V_1 \times V_2 \rightarrow \mathbb{F}_2$ such that

$$\langle x_1 + x_2, y_1 + y_2 \rangle = \langle x_1, y_1 \rangle + \langle x_1, y_2 \rangle + \langle x_2, y_1 \rangle + \langle x_2, y_2 \rangle$$

for every $x_1, x_2 \in V_1$ and $y_1, y_2 \in V_2$. A bilinear form is *non-degenerate* if for every $0 \neq x \in V_1$ there exists $y \in V_2$ such that $\langle x, y \rangle = 1$, and for every $0 \neq y \in V_2$ there exists $x \in V_1$ such that $\langle x, y \rangle = 1$.

When possible, we will work with the standard bilinear form $\langle \rangle_{std} : \mathbb{F}_2^n \times \mathbb{F}_2^n \rightarrow \mathbb{F}_2$ given by $\langle x, y \rangle_{std} = \sum_{j=1}^n x_j y_j$ instead of arbitrary bilinear forms. The next claim allows us to move to this standard bilinear form.

► **Claim 120.** Let $\langle \rangle : V_1 \times V_2 \rightarrow \mathbb{F}_2$ be a non-degenerate bilinear form. Then $\dim(V_1) = \dim(V_2) =: m$ and there exists isomorphisms $\psi_i : V_i \rightarrow \mathbb{F}_2^m$ such that $\langle x, y \rangle = \langle \psi_1(x), \psi_2(y) \rangle_{std}$.

From now on, we write $\langle \rangle$ instead of $\langle \rangle_{std}$ when it is clear from context; Claim 120 justifies the abuse of notation.

Proof sketch of Claim 120. To see that $\dim(V_1) \leq \dim(V_2)$, note that from non-degeneracy the function $\pi : V_1 \rightarrow V_2^*$ that sends $x \mapsto \langle x, \cdot \rangle$ is an injective linear function. Swapping V_1 and V_2 shows that the dimensions are equal.

We observe that for every basis $f_1, \dots, f_m$ of V_1 there exists a *dual basis* $h_1, \dots, h_m$ of V_2 such that $\langle f_i, h_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$. As for the isomorphisms, we fix some basis $f_1, \dots, f_m$ of V_1

and its dual basis $h_1, \dots, h_m \in V_2$. Let $e_1, \dots, e_m \in \mathbb{F}_2^m$ be the standard basis. It is easy to verify that the pair of isomorphisms ψ_1 that sends f_i to e_i , and ψ_2 that sends h_i to e_i satisfy $\langle x, y \rangle = \langle \psi_1(x), \psi_2(y) \rangle$. $\triangleleft$

For a fixed bilinear form and a subspace $U \subseteq V_1$ the *orthogonal subspace* is the subspace

$$U^\perp = \{y \in V_2 \mid \forall x \in U, \langle x, y \rangle = 0\} \subseteq V_2,$$

and for a subspace $U \subseteq V_2$, one can analogously define $U^\perp \subseteq V_1$.

► **Definition 121** (Derived bilinear form). *Let $\langle \cdot \rangle : V_1 \times V_2 \rightarrow \mathbb{F}_2$ be a non-degenerate bilinear form and let $U \subseteq V_1$ be a subspace. The derived bilinear form $\langle \cdot \rangle_U : V_1/U \times U^\perp \rightarrow \mathbb{F}_2$ is a bilinear form given by*

$$\langle x + U, v \rangle_U = \langle x, v \rangle$$

for any $x + U \in V_1/U$ and $v \in U^\perp$.

It is easy to verify that $\langle x + U, v \rangle_U$ is well defined. That is, it does not depend on the representative $x \in x + U$. It is also straightforward to verify that this derived bilinear form is non-degenerate.

Most of the time we will abuse notation and write $\langle x + U, v \rangle$ instead of $\langle x + U, v \rangle_U$ when U is clear from the context.

7.2 The induced Grassmann poset - definition and running examples

In this subsection we construct a type of Grassmann posets called *induced Grassmann posets*. This construction generalizes the matrix Grassmann poset [50], its sparsified version in Definition 40 and the Johnson Grassmann poset in Definition 58.

Let $s : \mathbb{F}_2^d \setminus \{0\} \rightarrow \mathbb{F}_2^n$ be a function, not necessarily linear. The *induced Hadamard encoding* of s is the linear map $\widehat{s} : \mathbb{F}_2^d \rightarrow \mathbb{F}_2^n$ where

$$\widehat{s}(x) := \sum_{v \in \mathbb{F}_2^d} \langle x, v \rangle s(v) = \sum_{v \in \mathbb{F}_2^d : \langle x, v \rangle = 1} s(v).$$

We note that the inner product between any x and the 0 vector is always 0. Thus we just need to define these functions over $s : \mathbb{F}_2^d \setminus \{0\} \rightarrow \mathbb{F}_2^n$.

We call $\widehat{s}$ the induced Hadamard encoding, because if $n = 2^d - 1$, we index the coordinates of $\mathbb{F}_2^n$ by the non-zero vectors of $\mathbb{F}_2^d$, and set $s : \mathbb{F}_2^d \setminus \{0\} \rightarrow \mathbb{F}_2^n$ to be the function that sends $v \in \mathbb{F}_2^d \setminus \{0\}$ to the standard basis vector that is 1 on the v -th coordinate and zero elsewhere. Then $\widehat{s}$ is the (usual) Hadamard encoding and its image is the Hadamard code.

► **Definition 122** (Induced Grassmann poset). *Let $d, n \geq 0$ be two integers such that $n \geq 2^d - 1$. Let $\mathcal{S}$ be a set of (non-linear) functions such that every $s \in \mathcal{S}$ is injective and the set $\text{Im}(s) := \{s(v) \mid v \in \mathbb{F}_2^d \setminus \{0\}\}$ is independent. The $\mathcal{S}$ -induced Grassmann poset is the d -dimensional Grassmann poset $Y^{\mathcal{S}}$ such that $Y^{\mathcal{S}}(d) = \{\text{Im}(\widehat{s}) \mid s \in \mathcal{S}\}$.*

To be explicit, for $i < d$, an i -dimensional subspace W is in $Y^{\mathcal{S}}(i)$ if there exists some $W' \in Y^{\mathcal{S}}(d)$ containing W . Equivalently, this is if and only if there exist a $s \in \mathcal{S}$ and a subspace $U \subseteq \mathbb{F}_2^d$ such that $\widehat{s}|_U = W$. For $i = 1$ this says that $v \in Y^{\mathcal{S}}(1)$ if and only if $v = \widehat{s}(x)$ for some $s \in \mathcal{S}$ and non-zero $x \in \mathbb{F}_2^d$.

We note that if the vectors in the image of s are independent, then $\dim(\text{Im}(\widehat{s})) = d$. We also note that different s, s' may still result in the same spaces $\text{Im}(\widehat{s}) = \text{Im}(\widehat{s'})$.

Although this definition seems abstract, hopefully, the following examples, which were already introduced in the previous sections, will convince the reader that this definition is natural.

► **Example 123** (The complete Grassmann). Let d and $n \geq 2^d - 1$ be positive integers. Let $\mathcal{S}$ be all injective functions whose image is independent. Then $Y^{\mathcal{S}} = \mathcal{G}(F_2^n)^{\leq d}$ is the d -skeleton of the complete Grassmann of $\mathbb{F}_2^n$. To see this let us show that every d -dimensional subspace W has some $s \in \mathcal{S}$ such that $\text{Im}(\widehat{s}) = W$. Let $s' \in \mathcal{S}$ be an arbitrary function and let $W' = \text{Im}(\widehat{s'})$. Find some $A \in GL_n(\mathbb{F}_2)$ that maps W' to W and set $s(x) := A(s'(x))$. It is easy to verify that indeed $s \in \mathcal{S}$ and that $\text{Im}(\widehat{s}) = W$.

► **Example 124** (Johnson Grassmann poset). Let d and $n \geq 2^d - 1$ be positive integers. Let $E = \{e_i\}_{i=1}^n$ be the standard basis for $\mathbb{F}_2^n$. Let $\mathcal{S}_J$ be the set of injective functions from $\mathbb{F}_2^d$ to E . For this $\mathcal{S}_J$, the poset $Y^{\mathcal{S}_J}$ is the Johnson Grassmann poset.

First note that a vector from $\mathbb{F}_2^n$ is in $Y^{\mathcal{S}_J}(1)$ if and only if its Hamming weight is 2^{d-1} : every such vector w is equal to $w = \widehat{s}(x) = \sum_{v \in \mathbb{F}_2^d} \langle x, v \rangle s(v)$ for some $s \in \mathcal{S}_J$ and $x \neq 0$. The $s(v)$'s are distinct standard basis vectors, and there are exactly 2^{d-1} non-zero $\langle x, v \rangle$ in the sum. Therefore the weight of w is 2^{d-1} .

The d -dimensional subspaces are the Hadamard encoding on $2^d - 1$ coordinates inside n . Let us be more precise. Fix a map s and for simplicity assume that $\text{Im}(s) = \{e_1, e_2, \dots, e_{2^d-1}\}$. Let $W = \text{span}(\text{Im}(s))$. By indexing W with the coordinates of $v \in \mathbb{F}_2^d \setminus \{0\}$, then $\widehat{s}(x) = (\langle x, v \rangle)_{v \in \mathbb{F}_2^d \setminus \{0\}} \in W$, which is the Hadamard encoding of x .

The basification of this poset is an alternative description of the links of Johnson complexes.

► **Example 125** (Matrix poset [50]). Let d and $n \geq 2^d - 1$. Recall that $\mathcal{M}(1)$ is the set of all rank one $n \times n$ matrices over $\mathbb{F}_2$. Let $\mathcal{S}_{\mathcal{M}}$ be the set of functions $s : \mathbb{F}_2^d \rightarrow \mathcal{M}(1)$ such that the matrix $M_s = \left(\bigoplus_{v \in \mathbb{F}_2^d \setminus \{0\}} s(v) \right)$ is a direct sum, or in other words $\text{rank}(M_s) = 2^d - 1$. One can see directly that such s 's are the admissible functions of Definition 39 for the special case where the ambient poset is complete Grassmann $\mathcal{G}(\mathbb{F}_2^n)$. In this case, $Y^{\mathcal{S}_{\mathcal{M}}} = \mathcal{SP}(\mathcal{G}(\mathbb{F}_2^n))$.²⁰

To draw the analogy to the Johnson, let us illustrate that indeed a matrix A is a 1-dimensional subspace in the $\mathcal{S}_{\mathcal{M}}$ -induced Grassmann poset $Y^{\mathcal{S}_{\mathcal{M}}}$ if and only if $\text{rank}(A) = 2^{d-1}$. This follows the same logic as in the Johnson case. If $A \in Y^{\mathcal{S}_{\mathcal{M}}}(1)$ then $A = \widehat{s}(x)$ for some $s \in \mathcal{S}_{\mathcal{M}}$ and $x \neq 0$. This means that A is the sum of 2^{d-1} distinct $s(v)$'s. As the matrix M_s is a direct sum, the sum of the $s(v)$'s is also a direct sum. This implies that $\text{rank}(A) = \sum_{x: \langle v, x \rangle = 1} \text{rank}(s(v)) = 2^{d-1}$. On the other hand, suppose A has rank 2^{d-1} and decomposition $A = \sum_{j=1}^{2^{d-1}} e_j \otimes f_j$. Take an arbitrary $x \in \mathbb{F}_2^d \setminus \{0\}$ and let $R = \{v \in \mathbb{F}_2^d \mid \langle v, x \rangle = 1\}$. As R also has size 2^{d-1} , we can find some $s \in \mathcal{S}$ that maps the elements in R to the $e_j \otimes f_j$ in the decomposition of A . For such an s , we have $\widehat{s}(x) = \sum_{v \in R} s(v) = A$, thus concluding $A \in Y^{\mathcal{S}}(1)$.

The construction in [50] used matrices over a larger field of characteristic 2 instead of $\mathbb{F}_2$ itself. See [50] for the details.

The next example is the one given in Definition 40.

► **Example 126** (Sparsified matrix poset). Let $d > 0$ and let Y_0 be any $2^d - 1$ -dimensional Grassmann poset over $\mathbb{F}_2^n$. We define a sparsification $\mathcal{S}_{Y_0} \subseteq \mathcal{S}_{\mathcal{M}}$.

$$\mathcal{S}_{Y_0} = \{s \in \mathcal{S}_{\mathcal{M}} \mid \text{row}(M_s), \text{col}(M_s) \in Y_0(2^d - 1)\}^{21}$$

²⁰There is a slight technical difference between this complex and the one in [50], that follows from Golowich's use of larger fields. See [50] for more details.

where $M_s = \left(\bigoplus_{v \in \mathbb{F}_2^d \setminus \{0\}} s(v) \right)$. The one dimensional spaces $M \in Y^{\mathcal{S}_{Y_0}}(1)$ are 2^{d-1} -rank matrices such that $\text{row}(M), \text{col}(M) \in Y_0(2^{d-1})$. The top-level spaces are $W \in Y^{\mathcal{S}_{\mathcal{M}}}(d)$, such that the sum of row spaces (resp. column spaces) is in $Y_0(2^d - 1)$.

Finally, we end this subsection with the comment that if one has a distribution μ over $\mathcal{S}$, this also defines a distribution over $Y^{\mathcal{S}}(d)$ in a natural way. That is, sampling $s \sim \mu$ and then outputting $\text{Im}(\widehat{s}) \in Y^{\mathcal{S}}(d)$. Unless otherwise stated, we will always take the uniform distribution over $Y^{\mathcal{S}}(d)$.

7.2.1 Subspaces of an induced Grassmann poset

Fix $i < d$. The subspaces $Y^{\mathcal{S}}(i)$ are defined by downward closure, but we can give an alternative description of $Y^{\mathcal{S}}(i)$ via an induced Grassmann poset with some $\mathcal{S}_i$ that is derived from $\mathcal{S}$. We begin by looking at our examples for intuition and then generalize.

In the examples above, the lower dimensional subspaces have some structure to them. For instance, one can verify that the i -dimensional subspaces in the Johnson Grassmann poset could be described as images of Hadamard encodings $\widehat{s}$ where the image of $s : \mathbb{F}_2^i \setminus \{0\} \rightarrow \mathbb{F}_2^n$ consists of vectors of Hamming weight 2^{d-i} such that the supports of every pair $s(v)$ and $s(v')$ are disjoint. This generalizes the case where $i = d$, and the image of $s \in \mathcal{S}_J$ are distinct vectors of weight 1.

In a similar manner, we saw in Claim 41 that i -dimensional subspaces of the Matrix Grassmann poset are images of induced Hadamard encodings

$$\text{Im}(\widehat{s}) = \left\{ \sum_{v \in \mathbb{F}_2^i} \langle x, v \rangle s(v) \mid x \in \mathbb{F}_2^i \right\},$$

where every $s(v)$ has rank 2^{d-i} and $M_s = \bigoplus_{v \in \mathbb{F}_2^i \setminus \{0\}} s(v)$ has rank $2^d - 2^{d-i}$. Again, this is a natural generalization of the case where $i = d$.

Back to the more general setup. Recall that every $W \in Y^{\mathcal{S}}(d)$ is the image of $\widehat{s}$ for some $s \in \mathcal{S}$. The map $\widehat{s}$ is always injective, therefore every i -dimensional subspace $W' \subseteq W$ is also the image $W' = \text{Im}(\widehat{s}|_U)$ of some i -dimensional subspace $U \subseteq \mathbb{F}_2^d$.

Let $V \subseteq \mathbb{F}_2^d$ be the $(d-i)$ -dimensional space such that $U = V^\perp$, that is, $U = \{x \in \mathbb{F}_2^d \mid \forall y \in V, \langle x, y \rangle = 0\}$. Let $s_V : (\mathbb{F}_2^d/V) \setminus \{0+V\} \rightarrow \mathbb{F}_2^n$ be $s_V(v+V) = \sum_{v' \in v+V} s(v')$. We use the bilinear form $\langle \rangle_V$ to define the function $\widehat{s}_V : U \rightarrow \mathbb{F}_2^n$ as the induced Hadamard encoding, i.e.

$$\widehat{s}_V(x) = \sum_{v+V \in \mathbb{F}_2^d/V} \langle v+V, x \rangle_V \cdot s_V(v+V).$$

Here we recall that $\langle v+V, x \rangle_V := \langle x, v \rangle$ for $x \in U$ (and we recall that this is independent of the choice of representative v).

Let

$$\mathcal{S}_i = \{s_V \mid s \in \mathcal{S}, V \subseteq \mathbb{F}_2^d, \dim(V) = d-i\}.$$

We will show that for every $s \in \mathcal{S}$, $\text{Im}(\widehat{s}|_U) = \text{Im}(\widehat{s}_V)$ and hence $\mathcal{S}_i$ induces $Y^{\mathcal{S}}(i)$.

▷ **Claim 127.** For every subspace $U \subseteq \mathbb{F}_2^d$, $V = U^\perp$, function $s : \mathbb{F}_2^d \rightarrow \mathbb{F}_2^n$, and $x \in U$, $\widehat{s}_V(x) = \widehat{s}(x)$. As a result, $(Y^{\mathcal{S}})^{\leq i} = Y^{\mathcal{S}_i}$.

Here there is some abuse of notation since the domain of $\widehat{s}_V$ is not $\mathbb{F}_2^i$, but some arbitrary i -dimensional space $U \subseteq \mathbb{F}_2^d$. However, this does not matter since we are taking the *image space* of $\widehat{s}_V$.

Before proving this claim, let us review our previous examples in light of this.

► **Example 128** (Johnson Grassmann poset). For the Johnson Grassmann poset, fix a $(d-i)$ -dimensional subspace $V \subseteq \mathbb{F}_2^d$. Then a function $s_V \in \mathcal{S}_{J,i}$ if and only if every value $s_V(v+V)$ has Hamming weight 2^{d-i} and for every $v+V \neq u+V$ the supports of $s_V(v+V)$ and $s_V(u+V)$ are mutually disjoint.

By Claim 120 we replace the functions in $\mathcal{S}_i$, with similar functions whose domain is $\mathbb{F}_2^i$, without changing the resulting poset. Thus another set of functions $\mathcal{S}'_i$ such that the induced Grassmann on $\mathcal{S}'_i$ is also equal to $(Y^{\mathcal{S}_J})^{\leq i}$ is the set of all functions $s : \mathbb{F}_2^i \setminus \{0\} \rightarrow \mathbb{F}_2^n$ such that the Hamming weight of every $s(v)$ is 2^{d-i} , and such that if $v \neq u$ then the supports of $s(v)$ and $s(u)$ are disjoint. These are exactly the admissible functions in Definition 69.

► **Example 129** (Matrix poset and sparsified matrix poset). For the matrix example above let us fix some $s \in \mathcal{S}_M$ and $(d-i)$ -dimensional subspace $V \subseteq \mathbb{F}_2^d$. The matrix $s_V(v+V)$ is a direct sum of 2^{d-i} rank 1 matrices. Thus $\text{rank}(s_V(v+V)) = 2^{d-i}$. The matrices $\{s_V(v+V)\}_{v+V}$ have the property that their sum $M_{s_V} = \bigoplus_{v+V \neq 0+V} s_V(v+V)$ is a direct sum, or equivalently that $\text{rank}(M_{s_V}) = 2^d - 2^{d-i}$.²²

Again by Claim 120, we can give an equivalent set of functions $\mathcal{S}_i$ – the level i *admissible* functions in Section 3. These are all functions $s : \mathbb{F}_2^i \setminus \{0\} \rightarrow \mathbb{F}_2^n$ such that $\text{rank}(s(v)) = 2^{d-i}$ and $\text{rank}(M_s) = 2^d - 2^{d-i}$ where $M_s = \sum_{v \in \mathbb{F}_2^i \setminus \{0\}} s(v)$.

We remark that the sparsified matrix induced Grassmann poset $\mathcal{S}_{Y_0}$ also admit a similar description of the i -spaces which was explained in Section 3.

Proof of Claim 127. Fix s and an i dimensional subspace U . Let $V = U^\perp$ be some $d-i$ -dimensional subspace in $\mathbb{F}_2^d$. Let $x \in U$. Note that while $\widehat{s}(x) = \sum_{v \neq 0} \langle x, v \rangle s(v)$, we can remove all elements $v \in V$ from the sum since x is orthogonal to them. We also recall that for every $v_1, v_2 \in v+V$, $\langle x, v_1 \rangle = \langle x, v_2 \rangle := \langle x, v+V \rangle$ so

$$\begin{aligned} \widehat{s}(x) &= \sum_{v \notin V} \langle x, v \rangle s(v) \\ &= \sum_{v+V \neq 0+V} \langle x, v+V \rangle \left(\sum_{v' \in v+V} s(v') \right) \\ &= \sum_{v+V \neq 0+V} \langle x, v+V \rangle s_V(v+V) = \widehat{s}_V(x). \end{aligned}$$

Thus in particular for every s , $\text{Im}(\widehat{s}|_U) = \text{Im}(\widehat{s}_V)$. As $\widehat{s}$ is injective, every subspace $W \in Y^{\mathcal{S}_i}(i)$ is the image $\text{Im}(\widehat{s}|_U)$ of some i -dimensional subspace $U \subseteq \mathbb{F}_2^d$. The claim follows. $\triangleleft$

7.3 A decomposition for the links

Our goal is to prove that links of some induced Grassmann posets decompose into tensor products of simple graphs, reproving and extending both Proposition 51 for the matrix Grassmann poset, Lemma 46 for the sparsified matrix poset, and Proposition 62 for the Johnson Grassmann poset.

²² Note that this is not the matrix M_s in Example 125, since the matrices $s(v)$ do not participate in this sum when $v \in V \setminus \{0\}$.

For this we need to zoom in to a less general class of induced Grassmann posets, which we call *rank-based induced Grassmann posets* with the *direct-meet property*. Let us define these.

7.3.1 Rank-based posets

Let $G \subseteq \mathbb{F}_2^n$ be a spanning set that does not contain 0. The rank of a vector v with respect to G , $\text{rank}_G(v)$, is the minimum number of elements in G necessary to write v as their sum.

► **Definition 130** (*k*-admissible spanning set). *We say G is k -admissible if for every v with $\text{rank}_G(v) \leq k$ there exists some u such that $\text{rank}_G(v + u) = \text{rank}_G(v) + \text{rank}_G(u) = k$.*

► **Example 131.**

1. For example in Johnson Grassmann posets, take $G_{\mathcal{J}} := \{e_1, e_2, \dots, e_n\}$ to be the standard basis. Then $\text{rank}_{G_{\mathcal{J}}}(v) = \text{wt}(v)$ is just its Hamming weight. This set is n -admissible.
2. For matrix Grassmann posets, taking $G_{\mathcal{M}} := \mathcal{M}(1)$ to be the set of rank one matrices inside $\mathbb{F}_2^{n \times n}$, then $\text{rank}_{G_{\mathcal{M}}}(M) = \text{rank}(M)$ is its usual matrix rank. This set is also n -admissible. This second example can be extended to G being the rank one k -tensors for $k > 2$ in $\mathbb{F}_2^{n \times n \times \dots \times n}$ as well.

The following notation of a direct sum generalizes the matrix poset direct sum.

► **Definition 132** (Abstract direct sum). *We say that $w_1 + w_2 + \dots + w_j$ is a direct sum and write $w_1 \oplus w_2 \oplus \dots \oplus w_j$, if $\text{rank}_G(w_1 + w_2 + \dots + w_j) = \sum_{\ell=1}^j \text{rank}_G(w_\ell)$. We also write $w_1 \leq w_2$ if $w_1 \oplus (w_2 - w_1)$.*

For $G_{\mathcal{J}}$ we have $w_1 \oplus w_2$ if and only if they are disjoint (as sets) and $w_1 \leq w_2$ if and only if $w_1 \subseteq w_2$. For $G_{\mathcal{M}}$ the definitions of $M_1 \oplus M_2$ and $M_1 \leq M_2$ coincide with the direct sum and matrix domination order defined in Section 2.4. It is easy to show that for any G the relation $\leq$ is always a partial order.

► **Definition 133** (Rank r -admissible functions). *Let d be an integer. Let G be a $(r(2^d - 1))$ -admissible set. A rank r admissible function is $s : \mathbb{F}_2^d \setminus \{0\} \rightarrow \mathbb{F}_2^n$ such that every $s(v)$ has $\text{rank}_G(s(v)) = r$, and such that $\bigoplus_{v \in \mathbb{F}_2^d \setminus \{0\}} s(v)$ is a direct sum. The rank r admissible set of functions $\mathcal{S}_G$ is the set containing all rank r admissible functions.*

We comment that this set $\mathcal{S}$ indeed gives rise to an induced Grassmann poset, because if the sum of the vectors in $\{s(v)\}_{v \in \mathbb{F}_2^d \setminus \{0\}}$ is a direct sum, then this set is independent. The reason is that if there is a nonempty subset $A \subseteq \{s(v)\}_{v \in \mathbb{F}_2^d \setminus \{0\}}$ such that $\sum_{v \in A} s(v) = 0$, then

$$\begin{aligned} \text{rank}_G \left(\sum_{v \in \mathbb{F}_2^d \setminus \{0\}} s(v) \right) &= \text{rank}_G \left(\sum_{v \in \mathbb{F}_2^d \setminus (A \cup \{0\})} s(v) \right) \\ &\leq \sum_{v \in \mathbb{F}_2^d \setminus (A \cup \{0\})} \text{rank}_G(s(v)) < \sum_{v \in \mathbb{F}_2^d \setminus \{0\}} \text{rank}_G(s(v)), \end{aligned}$$

which shows that in such a case the sum was not a direct sum in the first place.

For the rest of this section we will always assume there is some ambient $(r(2^d - 1))$ -admissible set G , and that $\mathcal{S} = \mathcal{S}_G$ with respect to this set.

► **Definition 134** (Rank-based induced Grassmann poset). *If $\mathcal{S}$ is the rank r admissible set with respect to some G , then $Y^{\mathcal{S}}$ is called a rank-based induced Grassmann poset.*

► **Example 135.** Both the Johnson Grassmann poset and the matrix Grassmann posets are rank-based induced Grassmann posets for $G_{\mathcal{J}}$ and $G_{\mathcal{M}}$ respectively.

► **Example 136 (Non example).** The d -skeleton of the complete Grassmann is *not* a rank-based induced Grassmann for any $d \geq 2$. To see this we observe that in a rank-based induced Grassmann poset $Y^{\mathcal{S}}$, every $w \in Y^{\mathcal{S}}(1)$ has $\text{rank}_G(w) = r2^{d-1} \geq 2$. This is because $\hat{s}(x)$ is always a direct sum of 2^{d-1} vectors of rank r . However, in the complete Grassmann $Y(1)$ contains all non-zero vectors, including the vectors in G which have rank one.

7.3.2 The direct-meet property

Unfortunately, this is still too general for a meaningful decomposition lemma. For this we also need to define another property of G which we will call the *direct-meet property*.

► **Definition 137 (Meet).** The Meet of a set of elements $T \subseteq \mathbb{F}_2^n$ is an element $\text{Meet}(T) \in \mathbb{F}_2^n$ such that

1. For every $t \in T$, $\text{Meet}(T) \leq t$ (in the abstract direct sum sense).
2. For every $v \in \mathbb{F}_2^n$, if $v \leq t$ for every $t \in T$, then $v \leq \text{Meet}(T)$.

This is a standard definition in poset theory. We note that $\text{Meet}(T)$ does not necessarily exist, but if it exists it is always unique.

► **Definition 138 (Direct-meet property).** We say that G has the direct-meet property if for every three elements, w_1, w_2, w_3 such that $w_1 \oplus w_2 \oplus w_3$, the meet of $\{w_1 \oplus w_2, w_1 \oplus w_3\}$ exists and is equal to w_1 .

We say $Y^{\mathcal{S}}$ has the *direct-meet property* if it is a rank-based induced complex, where the generating set G has the direct-meet property.

We note that this definition is not satisfied by all spanning sets G .

► **Example 139 (Non example).** Let $G = \{e_1, e_2, e_3, e_1 + e_4, e_2 + e_4, e_3 + e_4\} \subseteq \mathbb{F}_2^4$ where e_1, e_2, e_3, e_4 are the standard basis. Then G does not have the direct-meet property. To see this let $w_1 = e_1, w_2 = e_2, w_3 = e_3$. The triple $w_1 \oplus w_2 \oplus w_3$ is a direct sum. However, $e_1 + e_4 \leq w_1 + w_2 = e_1 + e_2 = (e_1 + e_4) + (e_2 + e_4)$ and $e_1 + e_4 \leq w_1 + w_3 = e_1 + e_3 = (e_1 + e_4) + (e_3 + e_4)$, but $e_1 + e_4 \not\leq e_1 = w_1$. Hence w_1 is not the meet of $w_1 \oplus w_2$ and $w_1 \oplus w_3$.

However, both the Johnson and the matrix sets have this property.

► **Example 140.** For the set $G_{\mathcal{J}} = \{e_1, e_2, \dots, e_n\}$ which defines $Y^{\mathcal{J}}$, the direct-meet property holds. Indeed, the meet of any set T is the intersection $\text{Meet}(T) = \bigcap_{t \in T} t$, where we treat the vectors as sets. In particular, the meet of $w_1 \oplus w_2$ and $w_1 \oplus w_3$ is w_1 .

► **Example 141.** A more interesting example is the low-rank decomposition case. Let G be the set of rank 1 matrices which defines $Y^{\mathcal{S}_{\mathcal{M}}}$. Contrary to the Johnson case, there exists a set $\{M_1, M_2\}$ with no meet. For example let e_1, e_2 be independent, let $M_1 = e_1 \otimes e_1 + e_2 \otimes e_2$ and let $M_2 = e_1 \otimes e_1 + e_2 \otimes (e_1 + e_2)$. Both $A = e_1 \otimes e_1$ and $B = e_2 \otimes (e_1 + e_2)$ are under M_1 and M_2 in the matrix domination order. However, $A \not\leq B$ and $B \not\leq A$ and therefore there is no meet for $\{M_1, M_2\}$. Nevertheless, the rank 1 matrices have the direct-meet property.

▷ **Claim 142.** The set $G_{\mathcal{M}}$ has the direct-meet property.

This claim essentially abstracts the observation in [50, Lemma 57]. We prove it in Section C.3.

Finally we state an elementary claim about the meet which we will use later on, extending the property from three elements to any finite number of elements.

▷ **Claim 143.** Let G be a set with the direct-meet property. Let $T = \{w_1, w_2, \dots, w_m\}$ be such that $\bigoplus_{j=1}^m w_j$ (with respect to G). Let $\hat{T} = \{v_1, v_2, \dots, v_r\}$ be linear combinations of T , i.e. $v_j = \bigoplus_{\ell \in I_j} w_\ell$ for some subsets $I_j \subseteq [m]$. Then

$$\text{Meet}(\hat{T}) = \bigoplus_{\ell \in \bigcap_{j=1}^r I_j} w_\ell.$$

The proof is just an induction on the size of $\hat{T}$. We omit it.

7.3.3 The general decomposition lemma

We move towards our decomposition lemma for links of a rank-based $Y^{\mathcal{S}}$ with the direct-meet property. Our components for the decomposition lemma are the following.

► **Definition 144** (z -domination graph). Let $z \in \mathbb{F}_2^n$ be of rank $4m$. The graph $\mathcal{T}^z$ has vertices

$$V = \{y \leq z \mid \text{rank}(y) = 2m\},$$

$$E = \{\{w_1 \oplus w_2, w_1 \oplus w_3\} \mid \text{rank}(w_i) = m, w_1 \oplus w_2 \oplus w_3 \leq z\}.$$

To sample an edge we sample a uniformly random decomposition $w_1 \oplus w_2 \oplus w_3 \oplus w_4 = z$, such that $\text{rank}(w_i) = m$ and output $\{y, y'\}$ such that $y = w_1 \oplus w_2, y' = w_1 \oplus w_3$. The set of edges is the support of the above distribution.

For $\mathcal{G}_{\mathcal{M}}$, this graph coincides with the subposet graphs introduced in Definition 43. For the Johnson Grassmann poset, this graph turns out to be the Johnson graph $\mathcal{T}^z \cong J(4m, 2m, m)$. This is because every vector in the graph has weight $2m$ and is contained in a z of weight $4m$. Having $y = w_1 \oplus w_2$ and $y' = w_1 \oplus w_3$ in this definition is equivalent to y and y' intersecting on half their coordinates.

We continue with our second component.

► **Definition 145** (z -link graph). Let $z \in \mathbb{F}_2^n$ be a vector whose rank is $r(2^d - 2^i)$ and let $m \leq r2^{i-2}$. The graph $\mathcal{H}_{z,m}$ has vertices

$$V = \{y \in \mathbb{F}_2^n \mid \text{rank}(y) = 2m, y \oplus z\},$$

$$E = \{\{w_1 \oplus w_2, w_1 \oplus w_3\} \mid \text{rank}(w_i) = m, z \oplus w_1 \oplus w_2 \oplus w_3\}.$$

We define the (weighted) edges according to the following sampling process. To sample an edge we sample a random $s \in \mathcal{S}$ such that $\bigoplus_{v \in \mathbb{F}_2^d} s(v) := t \geq z$. Then we uniformly sample $z' \leq t - z$ such that $\text{rank}(z') = 4m$ and output a random edge in $\mathcal{T}^{z'}$.

For $\mathcal{G}_{\mathcal{M}}$, this graph coincides with the disjoint graph introduced in Definition 44. For the Johnson Grassmann poset, this graph is also a Johnson graph, $\mathcal{H}_{z,m} \cong J(n - |z|, 2m, m)$. This is because every vector in the graph has weight $2m$ whose support is disjoint from z . The sampling process in this graph amounts to uniformly sampling a $4m$ -weighted set z' that is disjoint from z , and then taking a step inside its $\mathcal{T}^{z'}$. This is the same as sampling (uniformly) two subsets y, y' disjoint from z that intersect on half their coordinates.

The final component we need to state the decomposition lemma is this following claim.

▷ **Claim 146.** Let $Y^{\mathcal{S}}$ be a rank-based induced poset with the direct-meet property. Then for every $W \in Y^{\mathcal{S}}(i)$, and $s_1, s_2 \in \mathcal{S}_i$ such that $\text{Im}(\widehat{s}_1) = \text{Im}(\widehat{s}_2) = W$, it holds that $\text{Im}(s_1) = \text{Im}(s_2)$.

We will hence denote by $Q_W = \{z_1, z_2, \dots, z_{2^i-1}\}$ the image set of a given W . We also denote by $z_W = \bigoplus_{j=1}^{2^i-1} z_j$ the direct sum.

This claim is a consequence of the direct-meet property. We prove it later and for now, let us proceed to the decomposition lemma.

► **Lemma 147** (The decomposition lemma). *Let $W \in Y^{\mathcal{S}}(i)$ with $Q_W = \{z_j\}_{j=1}^{2^i-1}$ and $z_W = \bigoplus_{j=1}^{2^i} z_j$ as above. Then $Y_W \cong \mathcal{T}^{z_1} \otimes \mathcal{T}^{z_2} \otimes \dots \otimes \mathcal{T}^{z_{2^i-1}} \otimes \mathcal{H}_{z_W, 2^{i-2}}$.*

We observe that this lemma applied to $Y^{\mathcal{S}_{\mathcal{J}}}$ implies Proposition 62. Applied to $Y^{\mathcal{S}_{\mathcal{M}}}$ this lemma proves Proposition 51. We can even generalize this lemma slightly more to capture the sparsified matrix case in Lemma 46. See Section 7.4 for details.

In order to prove Lemma 147, we need to define the isomorphism. To do so, we give a combinatorial criterion for a vector y being in $Y_W^{\mathcal{S}}(1)$ in terms of its meets with the elements in Q_W .

▷ **Claim 148.** Let $W \in Y^{\mathcal{S}}(i)$ with $Q_W = \{z_1, z_2, \dots, z_{2^i-1}\}$ as above. Let $y \in \mathbb{F}_2^n$. Then y is in the link of W if and only if there exists unique $y_1, y_2, \dots, y_{2^i}$ with the following properties:

1. $y = \bigoplus_{j=1}^{2^i} y_j$ (that is, y is equal to the sum of all y_j 's and this sum is direct).
2. All y_j 's have equal rank.
3. $y_i \leq z_i$.

In this case the subspace $W' = W + \text{span}(y)$, has $Q_{W'} = \{z_j\}_{j=1}^{2^i-1} \cup \{z_j - y_j\}_{j=1}^{2^i}$.

Note that items 1 and 3 imply that the rank of y_j is half the rank of z_j .

Given Claim 148, we can define the isomorphism $\psi : Y_W \rightarrow \mathcal{T}^{z_1} \otimes \mathcal{T}^{z_2} \otimes \dots \otimes \mathcal{T}^{z_{2^i-1}} \otimes \mathcal{H}_{z_W, 2^{i-2}}$ by

$$\psi(y) = (y_1, y_2, \dots, y_{2^i}), \quad (7.1)$$

where the image consists of the unique y_j 's promised by the claim. This function is also a bijection by Claim 148.

To conclude the proof, we need to argue that an edge $\{y, y'\}$ is sampled by sampling the different components y_j, y'_j independently from one another. For this part, we basically need the following independence claim for the distribution we used when sampling $s \in \mathcal{S}$. Let P be the random variable obtained by sampling a function from $\mathcal{S}$. For a set $R \subseteq \mathbb{F}_2^d$, let $P|_R$ be the random variable obtained by restricting the function only to the vectors of R . Let $z_0 \in \mathbb{F}_2^n$. Then $P|_R$ and $P|_{\mathbb{F}_2^d \setminus R}$ are independent, if we condition on the event $\sum_{v \in R} P|_R(v) = z_0$.

▷ **Claim 149.** Let $\mathcal{S}$ be a rank-based poset. Let $R_1, R_2, \dots, R_m \subseteq \mathbb{F}_2^d \setminus \{0\}$ be a partition. Let $z_1, z_2, \dots, z_{m-1} \in \mathbb{F}_2^n$ be such that $\text{rank}(z_i) = |R_i| \cdot r$ and $z_1 \oplus z_2 \oplus \dots \oplus z_{m-1}$.

Consider the joint random variables $\{P_i = s|_{R_i}\}_{i=1}^m$. Then conditioned on the event that for every $i = 1, 2, \dots, m-1$

$$\sum_{v \in R_i} P_i(v) = z_i,$$

these random variables are independent.

We prove the lemma first, then go back to show the sub-claims.

Proof of Lemma 147. Let us denote by $\mathcal{G} = \mathcal{T}^{z_1} \otimes \mathcal{T}^{z_2} \otimes \dots \otimes \mathcal{T}^{z_{2^i-1}} \otimes \mathcal{H}_{z_W, r2^{i-2}}$. We define a map $\psi : Y_W^{\mathcal{S}}(1) \rightarrow \mathcal{G}$ mapping $y \mapsto (y_1, y_2, \dots, y_{2^i})$ as in (7.1). That is, the y_j 's are as in Claim 148. By Claim 148, this map is well defined for every y and sends link vertices into the vertices of $\mathcal{G}$. Moreover, by Claim 148 this is also a bijection, since the map

$$\psi^{-1}(y_1, y_2, \dots, y_{2^i}) := \bigoplus_{j=1}^{2^i} y_j$$

is indeed onto $\bar{Y}_W^{\mathcal{S}}(1)$ according to the claim.

Next we prove that a pair $\{y, y'\}$ is an edge in $Y_W^{\mathcal{S}}$ if and only if $\{\psi(y), \psi(y')\}$ is an edge in $\mathcal{G}$ (ignoring any edge weights for now). Recall that $\{y, y'\}$ is an edge in $Y_W^{\mathcal{S}}$ if and only if $y' \in Y_{W'}^{\mathcal{S}}(1)$ where $W' = W + \text{span}(y)$. By Claim 148, $Q_{W'} = \{y_j\}_{j=1}^{2^i} \cup \{z_j - y_j\}_{j=1}^{2^i-1}$. Let us re-index $t_j \in Q_{W'}$'s with

$$t_{j,b} = \begin{cases} y_j & b = 0 \\ z_j - y_j & b = 1 \end{cases}.$$

for $(j, b) \in ([2^i - 1] \times \{0, 1\}) \cup \{(2^i, 0)\}$. Moreover, by applying Claim 148 to W' we have that $y' \in Y_{W'}^{\mathcal{S}}(1)$ if and only if:

1. We can decompose $y' = \bigoplus_{(j,b) \in [2^i] \times \{0,1\}} y'_{j,b}$ such that the rank of every summand is equal.
2. For every $(j, b) \neq (2^i, 1)$, $y'_{j,b} \leq t_{j,b}$.
3. The final $y'_{2^i,1}$ is a direct sum with $z_{W'}$ (the sum of all elements in $Q_{W'}$).

So on the one hand, if $\psi(y), \psi(y')$ are neighbors in the tensor product $\mathcal{G}$, we can take $y'_{j,0} = \text{Meet}(y'_j, y_j)$ and $y'_{j,1} = \text{Meet}(y'_j, z_j - y_j)$ which will satisfy the above property (and is well defined by the properties of $\mathcal{T}^{z_j}$ and the direct-meet property). This shows that being an edge in $\mathcal{G}$ implies being an edge in $Y_W^{\mathcal{S}}$.

For the reverse direction, suppose there exists such a decomposition of y' . Then for every z_j the vectors $w_1 = y'_{j,0}, w_2 = y'_{j,1}, w_3 = t_{j,0} - y'_{j,0}$ and $w_4 = t_{j,1} - y'_{j,1}$ have that $w_1 \oplus w_2 = y'_j$ and $w_1 \oplus w_3 = y_j$. In addition they have equal rank and their direct sum is equal to z_j . Thus $\{y_j, y'_j\}$ is an edge in $\mathcal{T}^{z_j}$. For the final component, a similar argument applies: We decompose $w_1 = y'_{2^i,0}, w_2 = y_{2^i} - y'_{2^i,0}$ and $w_3 = y'_{2^i,1}$. All three vectors have rank 2^{i-2} and have a direct sum. Furthermore, $y_{2^i} = w_1 \oplus w_2$ and $y'_{2^i} = w_1 \oplus w_3$. As a result $\{y_{2^i}, y'_{2^i}\}$ is an edge in $\mathcal{H}_{z_W, r2^{i-2}}$. Thus $\{\psi(y), \psi(y')\}$ is an edge in $\mathcal{G}$.

Finally, we note that we can sample $\{y, y'\}$ by sampling every pair $\{y_j, y'_j\}$ independently. Indeed, when sampling an edge in the link of W , we are conditioning on sampling some $s \in \mathcal{S}$ so that for a fixed $(d-i)$ -dimensional subspace V_0 , the image of $s_{V_0} : \mathbb{F}_2^d/V_0 \setminus \{0\} \rightarrow \mathbb{F}_2^n$ is fixed. For every $j \leq 2^i - 1$, there exists some $v + V_0$ such that $\{y_j, y'_j\}$ depend only on the values of $v' \in V_0$ (this is the $v + V_0$ such that $y_j, y'_j \subseteq s_{V_0}(v + V_0)$). For the final component $\{y_{2^i}, y'_{2^i}\}$, this edge only depends on s restricted to $(0 + V_0) \setminus \{0\}$. By Claim 149, the values of s on the respective parts in each coset $v + V_0$ are independent once all values of s_{V_0} have been fixed. We conclude that indeed ψ is an isomorphism. $\blacktriangleleft$

We conclude this subsection with the proofs of Claim 146, Claim 148 and Claim 149.

Proof of Claim 146. We will prove something even stronger. With the notation in the claim, we show that there exists some $A \in GL_i(\mathbb{F}_2)$ such that $s_2(x) = s_1(Ax)^{23}$. Let $s_1, s_2 \in \mathcal{S}_i$ be such that $Im(\widehat{s}_1) = Im(\widehat{s}_2) = W$ and such that without loss of generality the domain of s_1 and s_2 is $\mathbb{F}_2^i \setminus \{0\}$. Observe that by Claim 143, $Meet(\{\widehat{s}(y) \mid y \in \mathbb{F}_2^i, \langle x, y \rangle = 1\}) = s(x)$, since the subset on the left consists of direct sums of a bunch of $s(x')$, and $s(x)$ is the only common component in these sums of $\widehat{s}(y)$.

Now assume that $Im(\widehat{s}_2) = Im(\widehat{s}_1)$. This implies that there is an isomorphism B such that $\widehat{s}_1(x) = \widehat{s}_2(Bx)$. On the other hand, by the above

$$\begin{aligned} s_1(x) &= Meet(\{\widehat{s}_1(y) \mid y \in \mathbb{F}_2^i, \langle x, y \rangle = 1\}) \\ &= Meet(\{\widehat{s}_2(By) \mid y \in \mathbb{F}_2^i, \langle x, y \rangle = 1\}) \\ &= Meet(\{\widehat{s}_2(By) \mid By \in \mathbb{F}_2^i, \langle x, B^{-1}By \rangle = 1\}) \\ &\stackrel{y'=By}{=} Meet(\{\widehat{s}_2(y') \mid y' \in \mathbb{F}_2^i, \langle x, B^{-1}y' \rangle = 1\}) \\ &= Meet(\{\widehat{s}_2(y') \mid y' \in \mathbb{F}_2^i, \langle (B^{-1})^\top x, y' \rangle = 1\}) \\ &= s_2((B^{-1})^\top x). \end{aligned}$$

Set $A = (B^{-1})^\top$ and we got $s_1(x) = s_2(Ax)$. $\triangleleft$

We note that the above proof used the intersection patterns of the Hadamard code.

Proof of Claim 148. Let $s \in \mathcal{S}$ and $U' \subseteq \mathbb{F}_2^d$ be such that $Im(\widehat{s}|_{U'}) = W'$ and let $U \subseteq U'$ be such that $Im(\widehat{s}|_U) = W$. Write $V = V' \oplus \text{span}\{b\}$ such that $U' = V'^\perp$ and $U = V^\perp$. We recall that every $v + V \in \mathbb{F}_2^d/V$ can be partitioned into two cosets of V' , $v + V = v + V' \sqcup (v + b) + V'$ and therefore $s_V(v + V) = s_{V'}(v + V') + s_{V'}((v + b) + V')$. By the fact that $s_{V'} \in \mathcal{S}_i$ this sum is also direct.

We first prove the only if direction. Let $x \in U' \setminus U$ be such that $\widehat{s}(x) = y$. By Claim 127, it also holds that $\widehat{s_{V'}}(x) = y$. By the fact that $x \in U' \setminus U$, the inner product $\langle x, b \rangle = 1$ and therefore, $\langle x, v + V' \rangle \neq \langle x, (v + b) + V' \rangle$. This implies that for every $v_0 + V \neq 0 + V$, exactly one of $y_j \in \{s_{V'}(v_0 + V'), s_{V'}((v_0 + b) + V')\}$ satisfies $y_j \leq y$. Indeed:

$$\begin{aligned} y &= \widehat{s_{V'}}(x) \\ &= \sum_{v+V' \not\subseteq 0+V} \langle x, v + V' \rangle s_{V'}(v + V') \\ &= \sum_{v_0|v_0+V \neq 0+V} (\langle x, v_0 + V' \rangle s_{V'}(v_0 + V') + \langle x, (v_0 + b) + V' \rangle s_{V'}(v_0 + b + V')). \end{aligned}$$

Assume that $\langle x, v_0 + V' \rangle = 1$ and thus $\langle x, (v_0 + b) + V' \rangle = 0$ (the other case is similar). In this case, because this whole sum is a direct sum every non-zero component in the sum is dominated by y . In particular $s_{V'}(v_0 + V') \leq y$. Moreover, any component that does not appear in this sum is a direct sum with y , and in particular $s_{V'}(v_0 + b + V') \not\leq y$.

Thus we re-index $z_j = s_V(v + V)$ (arbitrarily) and set y_j to be the component where $y_j \leq y$ and $y_j \leq s_V(v + V) = z_j$. We also set $y_{2^i} = s_{V'}(b)$ and note that y_{2^i} also participates in the linear combination of $\widehat{s_{V'}}(x)$ because $\langle x, b \rangle = 1$. All y_j 's have rank 2^{d-i-1} and by the discussion above, we have that $y = \sum_{j=1}^{2^i} y_j$. Moreover, $y_j \leq z_j$ for every $j \leq 2^i - 1$ and y_{2^i} is a direct sum with the sum of z_j 's. Thus, $Q_{W'} = \{s_{V'}(v + V') \mid v + V' \neq V'\} =$

²³ Note that a priori s_1 and s_2 could have different domains, so technically A is an isomorphism between the domain of s_1 and that of s_2 .

$\{y_j\}_{j=1}^{2^i} \cup \{z_j - y_j\}_{j=1}^{2^i-1}$ by construction. Finally, we note that when the admissible functions come from a rank-based construction that has the direct-meet property, then $y_j = \text{Meet}(y, s_V(v_0 + V))$ for some $v_0 \in V$. Thus, the y_j 's are unique.

Conversely, let us assume that the three items hold and prove that $W' \in Y(i+1)$. Let $s \in \mathcal{S}_i$ be such that $\text{Im}(\widehat{s}) = W$. Recall that $\mathcal{S}_{i+1}$ consists of all functions s' whose image contains vectors of rank 2^{d-i} whose sum is a direct sum. Therefore, let us construct such a function $s' : \mathbb{F}_2^{i+1} \rightarrow \mathbb{F}_2^n$ satisfying $\text{Im}(\widehat{s}') = W'$. First, we re-index the elements y_j as y_v 's for every $v \in \mathbb{F}_2^i \setminus \{0\}$ (but keep the notation y_{2^i} as before) so that $y_v \leq s(v)$. For an input $(b, v) \in (\mathbb{F}_2 \times \mathbb{F}_2^i) \setminus \{0\}$,

$$s'(b, v) = \begin{cases} s(v) - y_v & b = 0, v \neq 0 \\ y_v & b = 1, v \neq 0 \\ y_{2^i} & b = 1, v = 0 \end{cases}.$$

It is easy to see that the image of s' is $\{y_j\}_{j=1}^{2^i} \cup \{z_j - y_j\}_{j=1}^{2^i-1}$, and these vectors form a direct sum and they all have rank 2^{d-i-1} . Thus we successfully construct a $s' \in \mathcal{S}_{i+1}$. Moreover, one readily verifies that $\widehat{s}'(1, \vec{0}) = \sum_{v \in \mathbb{F}_2^i} s'(1, v) = \sum_{j=1}^{2^i} y_j = y$ and that for $U'' = \{(0, v) \mid v \in \mathbb{F}_2^i\}$ we have that $\text{Im}(\widehat{s}'|_{U''}) = W$. Thus $\text{Im}(\widehat{s}') = W'$. $\triangleleft$

Proof of Claim 149. It suffices to prove that P_i is independent of $P_{[m] \setminus \{i\}}$ where $P_{[m] \setminus \{i\}}$ denotes the rest of the random variables. For any $i = 1, 2, \dots, m-1$, the distribution for P_i is uniform over all partial functions $s_0 : R_i \rightarrow \mathbb{F}_2^d$ satisfying (1) $\forall v \in R_i$ $\text{rank}(s_0(v)) = r$, and (2) $\bigoplus_{v \in R_i} s_0(v_i) = z_i$. This follows from the fact that for *any fixed conditioning* of the form $P_j = s|_{(\mathbb{F}_2^d \setminus \{0\}) \setminus R_i}$ where the image of P_j sum up to z_j for all $j \neq i$, all such functions s_0 have equal probability density.

For $i = m$, a similar argument applies. Conditioned on $P_1 = s|_{R_1}, P_2 = s|_{R_2}, \dots, P_{m-1} = s|_{R_{m-1}}$ (where the sum of each $s|_{R_i}$'s values equals z_i), the last partial function $s|_{R_m}$ is sampled by picking a uniformly random (ordered) set of rank r elements $t_1, t_2, \dots, t_{|R_m|}$ satisfying that $(\bigoplus_{j=1}^{m-1} z_j) \oplus (\bigoplus_{j=1}^{|R_m|} t_j)$. It is obvious that the distribution of P_m is independent of the specific functions assigned to $P_1, P_2, \dots, P_{m-1}$. $\triangleleft$

7.4 The distribution over admissible functions

We can extend the decomposition lemma above to a slightly more general setup. In this setup we sample $s \in \mathcal{S}$ according to a specified distribution instead of sampling it uniformly. Let $\mathcal{S}$ be a set of admissible functions. Let μ be a probability density function of $\mathcal{S}$. We say that μ is permutation invariant if for every $s \in \mathcal{S}$ and every permutation π of $\mathbb{F}_2^d \setminus \{0\}$, $\mu(s) = \mu(s \circ \pi)$. Such a μ induces the following measure over the d -dimensional subspaces $Y^{\mathcal{S}}(d)$ in the rank-based induced poset: first sample $s \sim \mu$, and then outputs $\text{Im}(\widehat{s})$. We call posets obtained by this process *measured rank-based posets*.

The sparsified matrix poset is an example of a measured rank-based posets, where the distribution μ of admissible functions is given by the following:

1. Sample $W_1, W_2 \in Y_0(2^d - 1)$ independently.
2. Sample s uniformly at random such that $\text{row}(\bigoplus_{v \in \mathbb{F}_2^d \setminus \{0\}} s(v)) = W_1$ and $\text{col}(\bigoplus_{v \in \mathbb{F}_2^d \setminus \{0\}} s(v)) = W_2$.

For general Y_0 , this distribution μ is nonuniform, and the resulting poset is a measured rank-based poset. Hence this more generalized definition captures the sparsified matrix poset in Definition 40.

For such posets we can prove a version of the decomposition lemma using the following variants of the graphs from Definition 145.

► **Definition 150** ($((z, \mu)$ -link graph). Let $z \in \mathbb{F}_2^n$ be of rank $2^d - 2^i$ and let $m \leq 2^{i-2}$. The graph $\mathcal{H}_z^\mu$ has vertices

$$V = \{y \in \mathbb{F}_2 \mid \text{rank}(y) = 2m, y \oplus z\},$$

$$E = \{\{w_1 \oplus w_2, w_1 \oplus w_3\} \in \mathbb{F}_2 \mid \text{rank}(w_i) = m, z \oplus w_1 \oplus w_2 \oplus w_3\}.$$

The edge distribution is given by: first sample a random vector $s \sim \mu$ conditioned on $t := \bigoplus_{v \in \mathbb{F}_2^d \setminus \{0\}} s(v) \geq z$. Then uniformly sample a $z' \leq t$ of rank $4m$ and finally output an edge in $\mathcal{T}^{z'}$ according to the distribution in that graph.

► **Lemma 151.** Let $Y^\mathcal{S}$ be a measured rank-based poset with the direct-meet property. Then for every $W \in Y^\mathcal{S}(i)$ with $Q_W = \{z_1, z_2, \dots, z_{2^i-1}\}$ and $z_W = \sum_{j=1}^{2^i-1} z_j$ defined as before,

$$Y_W^\mathcal{S} \cong \mathcal{T}^{z_1} \otimes \mathcal{T}^{z_2} \otimes \dots \otimes \mathcal{T}^{z_{2^i-1}} \otimes \mathcal{H}_{z_W, r2^{i-2}}^\mu.$$

The proof of this lemma follows the exact same argument in the proof of the decomposition lemma Lemma 147. Therefore we omit the proof for brevity.

References

- 1 Vedat Alev, Fernando Jeronimo, and Madhur Tulsiani. Approximating constraint satisfaction problems on high-dimensional expanders. In *Proc. 59th IEEE Symp. on Foundations of Computer Science*, pages 180–201, November 2019. doi:10.1109/FOCS.2019.00021.
- 2 Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992. doi:10.1002/RSA.3240030308.
- 3 Noga Alon and Yuval Roichman. Random cayley graphs and expanders. *Random Structures & Algorithms*, 5(2):271–284, 1994. doi:10.1002/RSA.3240050203.
- 4 Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence i: Modified log-sobolev inequalities for fractionally log-concave distributions and high-temperature ising models. *arXiv preprint*, 2021. arXiv:2106.04105.
- 5 Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials ii: high-dimensional walks and an fpras for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1–12, 2019. doi:10.1145/3313276.3316385.
- 6 Anurag Anshu, Nikolas P Breuckmann, and Chinmay Nirkhe. Nlts hamiltonians from good quantum codes. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1090–1096, 2023. doi:10.1145/3564246.3585114.
- 7 Sanjeev Arora and Madhu Sudan. Improved low degree testing and its applications. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 485–495, El Paso, Texas, 1997. doi:10.1145/258533.258642.
- 8 Mitali Bafna, Max Hopkins, Tali Kaufman, and Shachar Lovett. Hypercontractivity on high dimensional expanders. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 185–194, 2022. doi:10.1145/3519935.3520040.
- 9 Mitali Bafna, Noam Lifshitz, and Dor Minzer. Constant degree direct product testers with small soundness, 2024. arXiv:2402.00850.
- 10 Mitali Bafna and Dor Minzer. Characterizing direct product testing via coboundary expansion. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1978–1989, 2024. doi:10.1145/3618260.3649714.

- 11 Mitali Bafna, Dor Minzer, and Nikhil Vyas. Quasi-linear size pcps with small soundness from hdx. *arXiv preprint*, 2024. doi:10.48550/arXiv.2407.12762.
- 12 Cristina M Ballantine. Ramanujan type buildings. *Canadian Journal of Mathematics*, 52(6):1121–1148, 2000.
- 13 Imre Bárány. A generalization of carathéodory’s theorem. *Discrete Mathematics*, 40(2-3):141–152, 1982. doi:10.1016/0012-365X(82)90115-7.
- 14 Yonatan Bilu and Nathan Linial. Constructing expander graphs by 2-lifts and discrepancy vs. spectral gap. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 404–412. IEEE, 2004. doi:10.1109/F0CS.2004.19.
- 15 Endre Boros and Zoltán Füredi. The number of triangles covering the center of an n -set. *Geometriae Dedicata*, 17:69–77, 1984.
- 16 Donald I Cartwright, Patrick Solé, and Andrzej Żuk. Ramanujan geometries of type an. *Discrete mathematics*, 269(1-3):35–43, 2003.
- 17 Michael Chapman, Nathan Linial, and Peled Yuval. Expander graphs — both local and global. *Combinatorica*, 40:473–509, 2020. doi:10.1007/s00493-019-4127-8.
- 18 Michael Chapman and Alexander Lubotzky. Stability of homomorphisms, coverings and cocycles i: Equivalence, 2023. arXiv:2310.17474.
- 19 Michael Chapman and Alexander Lubotzky. Stability of homomorphisms, coverings and cocycles ii: Examples, applications and open problems, 2023. arXiv:2311.06706.
- 20 Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1548–1557. SIAM, 2021.
- 21 Fan Chung and Paul Horn. The spectral gap of a random subgraph of a graph. *Internet Mathematics*, 4(2-3):225–244, 2007. doi:10.1080/15427951.2007.10129296.
- 22 David Conlon. Hypergraph expanders from cayley graphs. *Israel Journal of Mathematics*, 233(1):49–65, 2019.
- 23 David Conlon, Jonathan Tidor, and Yufei Zhao. Hypergraph expanders of all uniformities from cayley graphs. *Proceedings of the London Mathematical Society*, 121(5):1311–1336, 2020.
- 24 Philippe Delsarte and Vladimir I. Levenshtein. Association schemes and coding theory. *IEEE Transactions on Information Theory*, 44(6):2477–2504, 1998. doi:10.1109/18.720545.
- 25 Yotam Dikstein. New high dimensional expanders from covers. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 826–838, 2023. doi:10.1145/3564246.3585183.
- 26 Yotam Dikstein and Irit Dinur. Agreement testing theorems on layered set systems. In *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science*, 2019.
- 27 Yotam Dikstein and Irit Dinur. Agreement theorems for high dimensional expanders in the small soundness regime: the role of covers. *CoRR*, abs/2308.09582, 2023. doi:10.48550/arXiv.2308.09582.
- 28 Yotam Dikstein and Irit Dinur. Coboundary and cosystolic expansion without dependence on dimension or degree. *CoRR*, abs/2304.01608, 2023. doi:10.48550/arXiv.2304.01608.
- 29 Yotam Dikstein and Irit Dinur. Swap cosystolic expansion. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1956–1966, 2024. doi:10.1145/3618260.3649780.
- 30 Yotam Dikstein, Irit Dinur, Yuval Filmus, and Prahladh Harsha. Boolean functions on high-dimensional expanders. *arXiv preprint arXiv:1804.08155*, 2018. arXiv:1804.08155.
- 31 Yotam Dikstein, Irit Dinur, and Alexander Lubotzky. Low acceptance agreement tests via bounded-degree symplectic hdxs, 2024. URL: <https://eccc.weizmann.ac.il/report/2024/019/>.
- 32 Yotam Dikstein and Max Hopkins. Chernoff bounds and reverse hypercontractivity on hdx. *arXiv preprint arXiv:2404.10961*, 2024. doi:10.48550/arXiv.2404.10961.
- 33 Irit Dinur. The pcp theorem by gap amplification. *Journal of the ACM (JACM)*, 54(3):12-es, 2007.

- 34 Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Good locally testable codes. *arXiv preprint arXiv:2207.11929*, 2022. doi:10.48550/arXiv.2207.11929.
- 35 Irit Dinur, Yuval Filmus, Prahladh Harsha, and Madhur Tulsiani. Explicit sos lower bounds from high-dimensional expanders. In *12th Innovations in Theoretical Computer Science (ITCS 2021)*, 2021. arXiv:2009.05218.
- 36 Irit Dinur and Elazar Goldenberg. Locally testing direct products in the low error range. In *Proc. 49th IEEE Symp. on Foundations of Computer Science*, 2008.
- 37 Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma. List-decoding with double samplers. *SIAM journal on computing*, 50(2):301–349, 2021. doi:10.1137/19M1276650.
- 38 Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 974–985. IEEE, 2017. doi:10.1109/FOCS.2017.94.
- 39 Irit Dinur, Siqi Liu, and Rachel Yun Zhang. New codes on high dimensional expanders. *arXiv preprint arXiv:2308.15563*, 2023. doi:10.48550/arXiv.2308.15563.
- 40 Irit Dinur and Inbal Livni Navon. Exponentially small soundness for the direct product z-test. In *32nd Computational Complexity Conference, CCC 2017, July 6-9, 2017, Riga, Latvia*, pages 29:1–29:50, 2017. doi:10.4230/LIPICS.CCC.2017.29.
- 41 Irit Dinur and Roy Meshulam. Near coverings and cosystolic expansion. *Archiv der Mathematik*, 118(5):549–561, May 2022. doi:10.1007/s00013-022-01720-6.
- 42 Dominic Dotterrer and Matthew Kahle. Coboundary expanders. *Journal of Topology and Analysis*, 4(04):499–514, 2012.
- 43 Dominic Dotterrer, Tali Kaufman, and Uli Wagner. On expansion and topological overlap. *Geometriae Dedicata*, 195:307–317, 2018.
- 44 Shai Evra and Tali Kaufman. Bounded degree cosystolic expanders of every dimension. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 36–48, 2016. doi:10.1145/2897518.2897543.
- 45 Ofer Gabber and Zvi Galil. Explicit constructions of linear-sized superconcentrators. *Journal of Computer and System Sciences*, 22(3):407–420, 1981. doi:10.1016/0022-0000(81)90040-4.
- 46 Jason Gaitonde, Max Hopkins, Tali Kaufman, Shachar Lovett, and Ruizhe Zhang. Eigenstriping, spectral decay, and edge-expansion on posets. *arXiv preprint arXiv:2205.00644*, 2022. doi:10.48550/arXiv.2205.00644.
- 47 Howard Garland. p -adic curvature and the cohomology of discrete subgroups of p -adic groups. *Ann. of Math.*, 97(3):375–423, 1973. doi:10.2307/1970829.
- 48 Oded Goldreich and Shmuel Safra. A combinatorial consistency lemma with application to proving the PCP theorem. In *RANDOM: International Workshop on Randomization and Approximation Techniques in Computer Science*. LNCS, 1997.
- 49 Louis Golowich. Improved Product-Based High-Dimensional Expanders. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2021)*, volume 207, 2021. doi:10.4230/LIPICS.APPROX/RANDOM.2021.38.
- 50 Louis Golowich. From grassmannian to simplicial high-dimensional expanders. *arXiv preprint arXiv:2305.02512*, 2023. doi:10.48550/arXiv.2305.02512.
- 51 Roy Gotlib and Tali Kaufman. List agreement expansion from coboundary expansion, 2022. arXiv:2210.15714.
- 52 M. Gromov. Singularities, expanders and topology of maps. part 2: from combinatorics to topology via algebraic isoperimetry. *Geom. Funct. Anal.*, 20:416–526, 2010. doi:10.1007/s00039-010-0073-8.
- 53 Anna Gundert and Uli Wagner. On eigenvalues of random complexes. *Israel Journal of Mathematics*, 216:545–582, 2016.
- 54 Tom Gur, Noam Lifshitz, and Siqi Liu. Hypercontractivity on high dimensional expanders. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 176–184, 2022. doi:10.1145/3519935.3520004.

- 55 Christopher Hoffman, Matthew Kahle, and Elliot Paquette. The threshold for integer homology in random d -complexes. *Discrete & Computational Geometry*, 57:810–823, 2017. doi:10.1007/S00454-017-9863-1.
- 56 Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.
- 57 Max Hopkins and Ting-Chun Lin. Explicit lower bounds against $\omega(n)$ -rounds of sum-of-squares. *arXiv preprint arXiv:2204.11469*, 2022. doi:10.48550/arXiv.2204.11469.
- 58 Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query PCPs. *SIAM J. Comput.*, 41(6):1722–1768, 2012. doi:10.1137/09077299X.
- 59 Fernando Granha Jeronimo, Shashank Srivastava, and Madhur Tulsiani. Near-linear time decoding of ta-shma’s codes via splittable regularity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1527–1536, 2021. doi:10.1145/3406325.3451126.
- 60 Tali Kaufman, David Kazhdan, and Alexander Lubotzky. Isoperimetric inequalities for ramanujan complexes and topological expanders. *Geometric and Functional Analysis*, 26(1):250–287, 2016.
- 61 Tali Kaufman and Alexander Lubotzky. High dimensional expanders and property testing. In *Innovations in Theoretical Computer Science, ITCS’14, Princeton, NJ, USA, January 12-14, 2014*, pages 501–506, 2014. doi:10.1145/2554797.2554842.
- 62 Tali Kaufman and David Mass. High dimensional random walks and colorful expansion. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017.
- 63 Tali Kaufman and David Mass. Good distance lattices from high dimensional expanders. (manuscript), 2018. arXiv:1803.02849.
- 64 Tali Kaufman and David Mass. Unique-neighbor-like expansion and group-independent cosystolic expansion. In *32nd International Symposium on Algorithms and Computation (ISAAC 2021)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021.
- 65 Tali Kaufman and David Mass. Double balanced sets in high dimensional expanders. *arXiv preprint arXiv:2211.09485*, 2022. doi:10.48550/arXiv.2211.09485.
- 66 Tali Kaufman and Izhar Oppenheim. Construction of new local spectral high dimensional expanders. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 773–786, 2018. doi:10.1145/3188745.3188782.
- 67 Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Combinatorica*, 40:245–281, 2020. doi:10.1007/S00493-019-3847-0.
- 68 Tali Kaufman and Izhar Oppenheim. Coboundary and cosystolic expansion from strong symmetry. In *48th International Colloquium on Automata, Languages, and Programming, ICALP 2021, July 12-16, 2021, Glasgow, Scotland (Virtual Conference)*, volume 198 of *LIPIcs*, pages 84:1–84:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. doi:10.4230/LIPIcs.ICALP.2021.84.
- 69 Tali Kaufman and Ran J Tessler. Garland’s technique for posets and high dimensional grassmannian expanders. *arXiv preprint*, 2021. arXiv:2101.12621.
- 70 Mikhail Koshchelev. Spectrum of johnson graphs. *Discrete Mathematics*, 346(3):113262, 2023. doi:10.1016/J.DISC.2022.113262.
- 71 Dmitry N Kozlov and Roy Meshulam. Quantitative aspects of acyclicity. *Research in the Mathematical Sciences*, 6(4):1–32, 2019.
- 72 Wen-Ching Winnie Li. Ramanujan hypergraphs. *Geometric & Functional Analysis GAFA*, 14:380–399, 2004.
- 73 Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. *Combinatorica*, 26(4):475–487, 2006. doi:10.1007/S00493-006-0027-9.
- 74 Siqi Liu, Sidhanth Mohanty, Tselil Schramm, and Elizabeth Yang. Local and global expansion in random geometric graphs. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023*, pages 817–825, New York, NY, USA, 2023. Association for Computing Machinery. doi:10.1145/3564246.3585106.

- 75 Siqi Liu, Sidhanth Mohanty, and Elizabeth Yang. High-dimensional expanders from expanders. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151, page 12. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020.
- 76 Alexander Lubotzky. Expander graphs in pure and applied mathematics. *Bulletin of the American Mathematical Society*, 49(1):113–162, 2012.
- 77 Alexander Lubotzky, Roy Meshulam, and Shahar Mozes. Expansion of building-like complexes. *Groups, Geometry, and Dynamics*, 10(1):155–175, 2016.
- 78 Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. doi:10.1007/BF02126799.
- 79 Alexander Lubotzky, Beth Samuels, and Uzi Vishne. Explicit constructions of ramanujan complexes of type ad. *European Journal of Combinatorics*, 26(6):965–993, 2005. doi:10.1016/J.EJC.2004.06.007.
- 80 Roger C Lyndon, Paul E Schupp, RC Lyndon, and PE Schupp. *Combinatorial group theory*, volume 89. Springer, 1977.
- 81 Neal Madras and Dana Randall. Markov chain decomposition for convergence rate analysis. *Annals of Applied Probability*, pages 581–606, 2002.
- 82 Adam W. Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing families i: Bipartite ramanujan graphs of all degrees. *Annals of Mathematics*, 182:307–325, 2015. doi:10.4007/annals.2015.182.1.7.
- 83 Adam W Marcus, Daniel A Spielman, and Nikhil Srivastava. Interlacing families ii: Mixed characteristic polynomials and the Kadison-Singer problem. *Annals of Mathematics*, pages 327–350, 2015.
- 84 Grigorii Aleksandrovich Margulis. Explicit constructions of concentrators. *Problemy Peredachi Informatsii*, 9(4):71–80, 1973.
- 85 Roy Meshulam and Nathan Wallach. Homological connectivity of random k-dimensional complexes. *Random Structures & Algorithms*, 34(3):408–417, 2009. doi:10.1002/RSA.20238.
- 86 Ryan O’Donnell and Kevin Pratt. High-dimensional expanders from chevalley groups. In Shachar Lovett, editor, *37th Computational Complexity Conference, CCC 2022, July 20-23, 2022, Philadelphia, PA, USA*, volume 234 of *LIPICs*, pages 18:1–18:26. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPICs.CCC.2022.18.
- 87 Izhar Oppenheim. Local spectral expansion approach to high dimensional expanders part i: Descent of spectral gaps. *Discrete & Computational Geometry*, 59(2):293–330, 2018. doi:10.1007/S00454-017-9948-X.
- 88 Pavel Panteleev and Gleb Kalachev. Asymptotically good quantum and locally testable classical LDPC codes. In Stefano Leonardi and Anupam Gupta, editors, *STOC ’22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 375–388. ACM, 2022. doi:10.1145/3519935.3520017.
- 89 Mark S Pinsker. On the complexity of a concentrator. In *7th International Teletraffic Conference*, volume 4, pages 1–318, 1973.
- 90 Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, June 1998. doi:10.1137/S0097539795280895.
- 91 Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability pcg characterization of np. In *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC ’97, pages 475–484, New York, NY, USA, 1997. ACM. doi:10.1145/258533.258641.
- 92 Omer Reingold. Undirected connectivity in log-space. *J. ACM*, 55(4), 2008. doi:10.1145/1391289.1391291.
- 93 Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 3–13. IEEE, 2000. doi:10.1109/SFCS.2000.892006.
- 94 Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM J. Comput.*, 25(2):252–271, 1996. doi:10.1137/S0097539793255151.

- 95 David Surowski. Covers of simplicial complexes and applications to geometry. *Geometriae Dedicata*, 16:35–62, April 1984. doi:10.1007/BF00147420.
- 96 Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251, 2017. doi:10.1145/3055399.3055408.
- 97 Egbert R Van Kampen. On some lemmas in the theory of groups. *American Journal of Mathematics*, 55(1):268–273, 1933.

A The decomposition lemma

Proof of Lemma 15. Let us denote by A_G, A_t the adjacency operators of G, G_t respectively. We also denote by $A_{\mathcal{H}}$ the *bipartite* adjacency operator of B_{τ} . That is, for a function $f : V \rightarrow \mathbb{R}$, $A_{\mathcal{H}}f : T \rightarrow \mathbb{R}$ is the function $A_{\mathcal{H}}f(t) = \mathbb{E}_{v \sim \mu_t} [f(v)]$. We note that by assumption, for every $f : V \rightarrow \mathbb{R}$ such that $\mathbb{E}_v [f(v)] = 0$, $\|A_{\mathcal{H}}f\|^2 \leq \lambda_2^2 \|f\|^2$.

Let $f : V \rightarrow \mathbb{R}$ be an eigenvector of A_G of norm 1 and denote its eigenvalue by μ . Then

$$\begin{aligned} |\mu| &= |\langle f, A_G f \rangle| = \mathbb{E}_{\{u,v\} \sim \mu_G} [f(u)f(v)] \\ &= \left| \mathbb{E}_{t \in \mu_T} \left[\mathbb{E}_{\{u,v\} \sim \mu_t} [f(u)f(v)] \right] \right| \\ &= \left| \mathbb{E}_{t \in \mu_T} [\langle f_t, A_t f_t \rangle] \right| \end{aligned}$$

where f_t is the restriction of f to the vertices of V_t . Let us denote $f_t = f_t^0 \cdot \mathbf{1} + f_t^\perp$ such that $f_t^0 \cdot \mathbf{1}$ is a constant and $f_t^\perp \perp f_t^0 \cdot \mathbf{1}$. Then by expansion,

$$\begin{aligned} &\leq \left| \mathbb{E}_{t \in \mu_T} [(f_t^0)^2 + \gamma \langle f_t^\perp, f_t^\perp \rangle] \right| \\ &= \left| \mathbb{E}_{t \in \mu_T} [(1-\gamma)(f_t^0)^2 + \gamma \langle f_t, f_t \rangle] \right| \\ &= (1-\gamma) \left| \mathbb{E}_{t \in \mu_T} [(f_t^0)^2] \right| + \gamma \left| \mathbb{E}_{v \sim V} [f(v)^2] \right| \\ &= (1-\gamma) \left| \mathbb{E}_{t \in \mu_T} [(f_t^0)^2] \right| + \gamma. \end{aligned}$$

Here first equality is algebra, the second is because $\mu_G(v)f(v)^2 = \sum_{v \in t} \mu_T(t)\mu_t(v)f(v)^2$ and the last equality is due to the fact that f has norm 1. Finally, we note that the constant $f_t^0 = \mathbb{E}_{v \sim \mu_t} [f(v)] = A_{\mathcal{H}}f(t)$. That is

$$= (1-\gamma)\|A_{\mathcal{H}}f\|^2 + \gamma \stackrel{(\|A_{\mathcal{H}}f\|^2 \leq \lambda_2^2)}{\leq} \gamma + (1-\gamma)\lambda_2^2. \quad \blacktriangleleft$$

B Intermediate subspaces of the expanding matrix Grassmann poset

In this appendix we prove that the intermediate subspaces in the Grassmann matrix poset come from level i -admissible functions.

▷ **Claim (Restatement of Claim 41).** Let Y' be as in Definition 40. Then

$$Y'(i) = \{Im(\hat{s}) \mid s \in \mathcal{S}_i\}.$$

For this proof we recall the preliminaries in Section 7 on quotient spaces and bilinear forms.

Proof of Claim 41. Fix d and $d' = \log d$. It is enough to prove that $W \in \{Im(\widehat{s}) \mid s \in \mathcal{S}_i\}$ if and only if there exists $W' \in Y'(d')$ such that $W' \supseteq W$ (i.e. if and only if $W \in Y'(i)$). We start with some $W \in Y'(i)$ and prove that $W \in \{Im(\widehat{s}) \mid s \in \mathcal{S}_i\}$. Let $W' \in Y'(d')$ be such that $w' \supseteq w$. Let $s \in \mathcal{S}_d$ be such that $Im(\widehat{s}) = W'$. The matrices in the image of s are linearly independent²⁴ since their sum has maximal rank. Hence $\widehat{s} : \mathbb{F}_2^{d'} \rightarrow W'$ is an isomorphism. From this there exists some i -dimensional subspace $U \subseteq \mathbb{F}_2^{d'}$ so that $W = Im(\widehat{s}|_U)$. We write $U = V^\perp$ for some $(d-i)$ -dimensional subspace V (with respect to the standard bilinear form). Let us define $s' : (\mathbb{F}_2^d/V) \setminus \{[0]\} \rightarrow \mathbb{F}_2^{n \times n}$ as

$$s'(v+V) = \sum_{v' \in v+V} s(v').$$

It is direct to check that s' is admissible, since every $\sum_{v' \in v+V} s(v')$ is a disjoint sum of $2^{d'-i}$ rank one matrices, and the sum $M_{s'} = \sum_{v+V \neq 0+V} s'(v+V)$ is also equal to

$$M_{s'} = \sum_{v \notin V} s(v)$$

which has rank $2^d - 2^{d-i}$ and whose row and column spaces are in Y_1 and Y_2 respectively, since it is contained in the row and column space of $M_s = \sum_{v \neq 0} s(v)$. We can also define $\widehat{s}' : U \rightarrow \mathbb{F}_2^{n \times n}$ as

$$\widehat{s}'(x) = \sum_{v+V \in (\mathbb{F}_2^d/V) \setminus \{[0]\}} \langle x, v+V \rangle s'(v+V).$$

Here the inner product is the derived inner product as in Section 2.2. We note that $\widehat{s}'(x) = \widehat{s}(x)$ because

$$\begin{aligned} \widehat{s}'(x) &= \sum_{v+V \in (\mathbb{F}_2^d/V) \setminus \{[0]\}} \langle x, v+V \rangle s'(v) \\ &= \sum_{v+V \in (\mathbb{F}_2^d/V) \setminus \{[0]\}} \langle x, v+V \rangle \sum_{v' \in v+V} s(v) \\ &\stackrel{(*)}{=} \sum_{v+V \in (\mathbb{F}_2^d/V) \setminus \{[0]\}} \sum_{v' \in v+V} \langle x, v' \rangle s(v') \\ &= \sum_{v \in \mathbb{F}_2^d \setminus \{0\}} \langle x, v \rangle s(v) = \widehat{s}(x) \end{aligned}$$

where $(*)$ is because the derived inner product $\langle x, v+V \rangle$ is equal to (the standard) inner product $\langle x, v' \rangle$ for every $v' \in v+V$. Even though the function s' is defined over U which is not equal to $\mathbb{F}_2^i$, and $\widehat{s}$ is defined through the derived bilinear form (not the standard one), by Claim 120 we can find an admissible s'' so that $Im(\widehat{s}'') = Im(\widehat{s})$. This proves that $w \in \{Im(\widehat{s}) \mid s \in \mathcal{S}_i\}$.

As for the converse direction, “reverse engineer” the proof above. Let $W \in \{Im(\widehat{s}) \mid s \in \mathcal{S}_i\}$, and we show that $W \in Y'(i)$. Let $s \in \mathcal{S}_i$ be such that $Im(\widehat{s}) = W$. Let $M_s = \sum_{v \in \mathbb{F}_2^i \setminus \{0\}} s(v)$ and let $W_1 \in (Y_1)_{\text{row}(M_s)}(d-1)$, $W_2 \in (Y_2)_{\text{col}(M_s)}(d-1)$. These spaces exist from the purity of Y_1 and Y_2 . We define $s' \in \mathcal{S}_{d'}$ as follows. For every $v \in \mathbb{F}_2^d \setminus \{0\}$ we (arbitrarily) find a decomposition of $s(v)$ into rank one matrices, and index them by $x \in \mathbb{F}_2^{d'-i}$.

²⁴ Here we say matrices are linearly independent if they are linear independent as vectors in $\mathbb{F}_2^{n^2}$.

That is, $s(v) = \sum_{x \in \mathbb{F}_2^{d'-i}} e_x \otimes f_x$. We set $s'(v, x) = e_x \otimes f_x$. For $v = 0 \in \mathbb{F}_2^i$ we define s' using W_1 and W_2 . We find space W'_1 and W'_2 such that $\text{row}(S) \oplus W'_j = W_j$ and ordered bases $e_1, e_2, \dots, e_{2^{d'-i-1}}$ to W_1 and $f_1, f_2, \dots, f_{2^{d'-i-1}}$ to W_2 . We (arbitrarily) re-index these bases as $\{e_x \mid x \in \mathbb{F}_2^{d'-i} \setminus 0\}$ and $\{f_x \mid x \in \mathbb{F}_2^{d'-i} \setminus 0\}$ and set $s'(0, x) = e_x \otimes f_x$. It is a direct definition chase to verify that s' is admissible, i.e. that $M_{s'} = \sum_{(v,x) \in \mathbb{F}_2^{d'} \setminus \{0\}} s'(v, x)$ has rank $2^{d'} - 1$ and its row and column spaces are equal to W_1 and W_2 respectively. Finally, the image of $\widehat{s}$ restricted to $U = \{(x, 0) \mid x \in \mathbb{F}_2^i\}$ is w since

$$\begin{aligned} \widehat{s}(x, 0) &= \sum_{(v,v') \in \mathbb{F}_2^i \times \mathbb{F}_2^{d'-i} \setminus \{(0,0)\}} \langle (x, 0), (v, v') \rangle s'(v, v') \\ &= \sum_{v \in \mathbb{F}_2^i \setminus \{0\}} \langle x, v \rangle \sum_{v' \in \mathbb{F}_2^{d'-i}} s'(v, v') \\ &= \sum_{v \in \mathbb{F}_2^i \setminus \{0\}} \langle x, v \rangle s(v) = \widehat{s}(x). \end{aligned}$$

Thus $W' = \text{Im}(\widehat{s}) \in Y'(d')$ and it contains $W = \text{Im}(\widehat{s}|_U)$, so $W \in Y'(i)$. $\triangleleft$

C Elementary properties of matrix poset

C.1 Domination is a partial order

Let us begin by showing that this relation is a partial order. We do so using this claim.

Proof of Claim 25. The implications of (4) $\Rightarrow$ (2) and (3) $\Rightarrow$ (2) are a direct consequence of the definitions. Let us see that (1) $\Rightarrow$ (2).

$$\begin{aligned} \text{rank}(A + B) &= \dim(\text{row}(A + B)) \\ &\leq \dim(\text{row}(A) + \text{row}(B)) \\ &= \dim(\text{row}(A)) + \dim(\text{row}(B)) - \dim(\text{row}(A) \cap \text{row}(B)) \\ &= \text{rank}(A) + \text{rank}(B) - \dim(\text{row}(A) \cap \text{row}(B)). \end{aligned}$$

This shows that if $\text{rank}(A + B) = \text{rank}(A) + \text{rank}(B)$ then $\text{row}(A) \cap \text{row}(B) = \{0\}$ (and similarly for the column space).

Finally, let us see that (2) $\Rightarrow$ (1), (3), (4). Assume that $\text{row}(A) \cap \text{row}(B), \text{col}(A) \cap \text{col}(B)$ are trivial. Let us write $r_1 = \text{rank}(A), r_2 = \text{rank}(B)$. Then we can write $A = \sum_{i=1}^{r_1} e_i \otimes f_i$ and $B = \sum_{i=1}^{r_2} g_i \otimes h_i$ and because the row and column spaces intersect trivially, we have that $\{e_i\} \cup \{g_i\}$ and $\{f_i\} \cup \{h_i\}$ are $r_1 + r_2$ sized sets that are independent, thus proving (4) (which implies (3)). This also shows that $\text{rank}(A + B) = \dim(\text{col}(A + B)) = \text{rank}(A) + \text{rank}(B)$, thus proving (1). $\triangleleft$

Proof of Claim 26. By Claim 25, $X \oplus Y$ if and only if $\text{col}(X) \cap \text{col}(Y)$ and $\text{row}(X) \cap \text{row}(Y)$ are $\{0\}$. So first of all note that if $C \oplus (B \oplus A)$ then $\text{row}(C) \cap \text{row}(B \oplus A) = \{0\}$ (and the same for the column space). By Claim 25 $\text{row}(B \oplus A) = \text{row}(B) \oplus \text{row}(A)$, so in particular $\text{row}(C) \cap \text{row}(B) = \{0\}$ (resp. $\text{col}(C) \cap \text{col}(B) = \{0\}$). Thus, $C \oplus B$.

Now let us show that $A \oplus (B \oplus C)$. By Claim 17, if $\text{row}(A) \oplus \text{row}(B)$ and $\text{row}(C) \oplus (\text{row}(A) \oplus \text{row}(B))$ then $\text{row}(A) \oplus (\text{row}(B) \oplus \text{row}(C))$ (resp. column spaces). Thus by Claim 25 $A \oplus (B \oplus C)$. $\triangleleft$

Proof of Claim 24. We need to verify reflexivity, anti-symmetry and transitivity. Let A, B, C be matrices.

1. Reflexivity is clear.
2. Anti-symmetry: assume that $A \leq B$ and $B \leq A$ this implies that $\text{rank}(B - A) = \text{rank}(A) - \text{rank}(B) = \text{rank}(B) - \text{rank}(A) = 0$ so $A = B$.
3. Transitivity: $C \leq B$ and $B \leq A$. Then by Claim 25, $\text{row}(A) = \text{row}(B) \oplus \text{row}(A - B)$. In addition this claim shows that $\text{row}(B - C) \subseteq \text{row}(B)$. Together this implies that $\text{row}(B - C)$ intersects trivially with $\text{row}(A - B)$. The same holds also for the column spaces. Thus by Claim 25 this implies that $\text{rank}(A - C) = \text{rank}(A - B) + \text{rank}(B - C)$, i.e. $A - C \geq A - B$. Thus,

$$\begin{aligned}
 \text{rank}(A) &\stackrel{(B \leq A)}{=} \text{rank}(B) + \text{rank}(A - B) \\
 &\stackrel{(C \leq B)}{=} (\text{rank}(C) + \text{rank}(C - B)) + \text{rank}(A - B) \\
 &\stackrel{(A - B \leq A - C)}{=} (\text{rank}(C) + \text{rank}(C - B)) + (\text{rank}(A - C) - \text{rank}(B - C)) \\
 &= \text{rank}(C) + \text{rank}(A - C). \quad \triangleleft
 \end{aligned}$$

► Remark 152. Recall the following definition from partially ordered set theory.

► Definition 153 (Rank function). Let $(A, \leq)$ be a poset. A rank function $\rho : \mathcal{M} \rightarrow \mathbb{N}$ is a function with the following properties.

1. If $A > B$ then $\rho(A) > \rho(B)$.
2. For every $A > B$ there exists $A \geq C > B$ so that $\rho(C) = \rho(B) + 1$.

A poset is called a graded poset if there exists a rank function for that poset.

The order of matrices is also a graded poset, where the matrix rank is the rank function. To see this, we note that if $A \leq B$, then $\text{rank}(A) = \text{rank}(B) + \text{rank}(A - B)$ so in particular $\text{rank}(A) \geq \text{rank}(B)$ and if $A \neq B$, $\text{rank}(A) > \text{rank}(B)$. Moreover, if $A > B$ then by the fourth item of Claim 25, we can write $A = B \oplus \left(\sum_{j=1}^t e_j \otimes f_j \right)$ for some $t > 0$. Taking $C = B \oplus (e_j \otimes f_j)$ concludes that the rank function (of matrices) is a rank function (of the graded poset).

C.2 Isomorphism of intervals

We begin with this claim.

- Claim 154. 1. Let $A \leq B, C$. Then $B \leq C$ if and only if $B - A \leq C - A$.
2. Let B, C be such that $B \oplus A$ and $C \oplus A$. Then $B \leq C$ if and only if $B \oplus A \leq C \oplus A$.

Proof. For the first item,

$$\begin{aligned}
 B &\leq C \\
 &\Leftrightarrow \text{rank}(C) = \text{rank}(B) + \text{rank}(C - B) \\
 &\Leftrightarrow \text{rank}(C) - \text{rank}(A) = \text{rank}(B) - \text{rank}(A) + \text{rank}(C - B) \\
 &\stackrel{(A \leq B, C)}{\Leftrightarrow} \text{rank}(C - A) = \text{rank}(B - A) + \text{rank}((C - A) - (B - A)) \\
 &\Leftrightarrow B - A \leq C - A.
 \end{aligned}$$

For the second item,

$$B \leq C$$

$$\Leftrightarrow \text{rank}(C) = \text{rank}(B) + \text{rank}(C - B)$$

$$\Leftrightarrow \text{rank}(C) + \text{rank}(A) = \text{rank}(B) + \text{rank}(A) + \text{rank}(C - B)$$

$$\Leftrightarrow \text{rank}(C \oplus A) = \text{rank}(B \oplus A) + \text{rank}((C \oplus A) - (B \oplus A))$$

$$\Leftrightarrow B \oplus A \leq C \oplus A.$$

◁

Proof of Proposition 28. By Claim 154 the bijection ψ maps A such that $A \leq M_2$ to $A' = A - M_1$ such that $A' \leq M_2 - M_1$. Therefore ψ is well defined as a function from $I(M_1, M_2)$ to $I(0, M_2 - M_1)$. Obviously ψ is injective (since it is injective over the set of all matrices). Moreover, if $M_1 \leq A_1 \leq A_2 \leq M_2$ then by Claim 154 the order is maintained, i.e. $A'_1 = A_1 - M_1 \leq A_2 - M_1 = A'_2$. The fact that $\text{rank}(A') = \text{rank}(A) - \text{rank}(M_1)$ follows from $M_1 \leq A$. Thus to conclude we need to show that ψ is surjective, or in other words, if $A' \leq M_2 - M_1$ then $A' + M_1 \leq M_2$. Let $A' \leq M_2 - M_1$. Then by Claim 25 $\text{row}(A') + \text{row}(M_2 - M_1 - A) = \text{row}(M_2 - M_1)$ and in particular $\text{row}(A') \subseteq \text{row}(M_2 - M_1)$ and similarly $\text{col}(A') \subseteq \text{col}(M_2 - M_1)$. By $M_1 \leq M_2$, we have that $\text{rank}(M_2) = \text{rank}(M_1) + \text{rank}(M_2 - M_1)$ and by Claim 25 this implies that $\text{row}(M_2 - M_1) \cap \text{row}(M_1) = \{0\}$ and similarly for the columns spaces. We conclude that $\text{row}(A') \cap \text{row}(M_1)$ and $\text{col}(A') \cap \text{col}(M_1)$ are trivial. By using Claim 25 we have that for $A = \psi^{-1}(A') = A' + M_1$,

$$\text{rank}(A) = \text{rank}(A') + \text{rank}(M_1)$$

or equivalently that $A \geq M_1$. Finally, we also have that $A \leq M_2$ by Claim 154 (as $A, M_2 \geq M_1$ and $A' = A - M_1 \leq M_2 - M_1$). Thus ψ is surjective since A is a preimage of A' inside $\{A \mid M_1 \leq A \leq M_2\}$. ◀

Proof of Claim 30. Fix A and note that $B \oplus A$ if and only if $A \leq A + B$. Thus ϕ is a bijection. The order preservation follows from the second item of Claim 154. ◁

C.3 The matrices have the direct-meet property

In this subsection we prove Claim 142.

Proof of Claim 142. The technical statement we shall use in the proof is the following: For any matrices $A, B \leq C$ such that $\text{row}(A) \subseteq \text{row}(B)$ and $\text{col}(A) \subseteq \text{col}(B)$, it holds that $A \leq B$.

Let us first show that this statement implies the direct meet property. Let M_1, M_2, M_3 be three matrices whose sum is direct. We observe that $\text{row}(M_1) = \text{row}(M_1 \oplus M_2) \cap \text{row}(M_1 \oplus M_3)$ ²⁵ and similarly $\text{col}(M_1) = \text{col}(M_1 \oplus M_2) \cap \text{col}(M_1 \oplus M_3)$. Consider any A such that $A \leq M_1 \oplus M_2$ and $A \leq M_1 \oplus M_3$. Observe that $\text{row}(A) \leq \text{row}(M_1 \oplus M_2) \cap \text{row}(M_1 \oplus M_3)$. Thus $\text{row}(A) \subseteq \text{row}(M_1)$. The same argument applies to the respective column spaces, and so by the technical statement above $A \leq M_1$. This implies that the set $G_{\mathcal{M}}$ has the direct meet property.

Now we prove the statement itself. Observe that $\text{row}(B - A) \subseteq \text{row}(B)$ and $\text{col}(B - A) \subseteq \text{col}(B)$. Moreover, $(C - B) \oplus B$ which means that $\text{row}(C - B)$ (respectively $\text{col}(C - B)$) intersects trivially with $\text{row}(B)$ (respectively $\text{col}(B)$). Thus $(C - B) \oplus (B - A)$. So we get

$$\text{rank}(C - A) = \text{rank}((C - B) + (B - A)) = \text{rank}(C - B) + \text{rank}(B - A).$$

²⁵ It is clear that $\text{row}(M_1) \subseteq \text{row}(M_1 \oplus M_2) \cap \text{row}(M_1 \oplus M_3)$. The other direction follows from a dimension count.

On the other hand, because $A, B \leq C$, the equality above can be rewritten as

$$\text{rank}(C) - \text{rank}(A) = \text{rank}(C) - \text{rank}(B) + \text{rank}(B - A)$$

or equivalently $\text{rank}(B) = \text{rank}(A) + \text{rank}(B - A)$ i.e. $A \leq B$. $\triangleleft$

D Properties of the graphs induced by the matrix poset

In this appendix we prove Lemma 47 and Claim 55 necessary to prove the expansion of the complexes constructed in Section 3. We also prove Claim 117 necessary for the coboundary expansion proven in Section 6. To do so, we need to take a technical detour and provide some further definitions that generalize the graphs in said statements. In Section 3 we defined some graphs that naturally arise from the relations $\leq$ and $\oplus$. To analyze them, we need to generalize these definitions.

► **Definition 155** (Relation graph). *Let $i < j \leq n$. The relation graph $\mathcal{R}(i, j) = (L, R, E)$ is the bipartite graph whose vertices are $L = \mathcal{M}(i), R = \mathcal{M}(j)$ and $\{A, B\} \in E$ if $A \leq B$.*

This is the analogous graph to the “down-up” graph in simplicial complexes, i.e. the graph whose vertices are $X(i)$ and $X(j)$ (for some simplicial complex X) and whose edges are all pairs $\{s, t\}$ such that $s \subseteq t$ (see e.g. [38]).

Another important graph is the following “sum graph”.

► **Definition 156** (Sum graph). *Let i, j, k be integers such that $k \leq i, j$ and $i + j - k \leq n$. The Sum graph $\mathcal{S}(i, j, k) = (L, R, E)$ is the bipartite graph whose vertices are $L = \mathcal{M}(i), R = \mathcal{M}(j)$ and $\{A, B\} \in E$ if there exists $C \in \mathcal{M}(k)$ such that $C \oplus (A - C) \oplus (B - C)$ are a direct sum.*

When $k = 0$ we denote this graph by $\mathcal{DS}(i, j) := \mathcal{S}(i, j, 0)$. In this case its definition simplifies: $A \sim B$ if and only if $A \oplus B$.

All the aforementioned graphs turn up when one analyzes the Grassmann graph defined in Section 3. However, the bulk of the analysis in Section 3 is on *links* of the Grassmann complex. There the graphs that appear are actually subgraphs of these graphs. Hence, we also make the following definitions of the graphs where we take $\mathcal{M}^U$ or $\mathcal{M}_{\cap D}$ as our underlying posets instead of all $\mathcal{M}$.

► **Definition 157** (Upper bounded relation graph). *Let $i < j \leq \ell \leq n$. Let $U \in \mathcal{M}(\ell)$. The graph $\mathcal{R}^U(i, j) = (L, R, E)$ is the bipartite graph whose vertices are $L = \mathcal{M}^U(i), R = \mathcal{M}^U(j)$ and $\{A, B\} \in E$ if $A \leq B$.*

Note that for every $U_1, U_2 \in \mathcal{M}(\ell)$, the graphs $\mathcal{R}^{U_1}(i, j), \mathcal{R}^{U_2}(i, j)$ are isomorphic. So when we do not care about the matrix we write $\mathcal{R}^\ell(i, j)$ instead of $\mathcal{R}^U(i, j)$.

The graphs $\mathcal{DS}^U(i, j)$ and $\mathcal{S}^U(i, j, k)$ are defined analogously as the subgraphs of matrices in $\mathcal{M}^U$:

► **Definition 158** (Upper bounded sum graph). *Let i, j, k, ℓ be integers such that $k \leq i, j$ and $i + j - k \leq \ell$. Let $U \in \mathcal{M}(\ell)$. The Sum graph $\mathcal{S}^U(i, j, k) = (L, R, E)$ is the bipartite graph whose vertices are $L = \mathcal{M}^U(i), R = \mathcal{M}^U(j)$ and $\{A, B\} \in E$ if there exists $C \in \mathcal{M}(k)$ such that $C \oplus (A - C) \oplus (B - C) \leq U$.*

In these cases we also write $\mathcal{DS}^\ell(i, j), \mathcal{S}^\ell(i, j, k)$ when we do not care about U .

The second type of localization is with respect to subspaces.

► **Definition 159** (Lower bounded local relation graph). Let i, j, ℓ be integers such that $i < j \leq n - \ell$. Let $D \in \mathcal{M}(\ell)$. The graph $\mathcal{R}_{\cap D}(i, j)$ is the subgraph of $\mathcal{R}(i, j)$ whose vertices are in $\mathcal{M}_{\cap D}$.

► **Definition 160** (Lower bounded sum graph). Let i, j, k, ℓ be integers such that $k \leq i, j$ and $i + j - k \leq n - \ell$. Let $D \in \mathcal{M}(\ell)$. The Sum graph $\mathcal{S}_{\cap D}(i, j, k) = (L, R, E)$ is the bipartite graph whose vertices are $L = \mathcal{M}_{\cap D}(i), R = \mathcal{M}_{\cap D}(j)$ and $\{A, B\} \in E$ if there exists $C \in \mathcal{M}(k)$ such that $D \oplus C \oplus (A - C) \oplus (B - C)$ (is a direct sum).

In the next section we prove upper bounds on the expansion of some of these graphs. In all the following claims $\gamma \in (0, 1)$ stands for universal constants which we don't calculate (and may change between claims). The main lemmas we need are these.

► **Lemma 161.** $\lambda_2(\mathcal{S}^\ell(i, j, k)) \leq \gamma^{-(\ell - j - \max(k, i - k))}$.

► **Lemma 162.** Let i, j, k, ℓ be such that $i \leq j$ and $2i \leq n - \ell$. Let $D \in \mathcal{M}(\ell)$. Then $\lambda_2(\mathcal{S}_{\cap D}(i, j, k)) \leq \gamma^{-(n - \ell - j - \max(k, i - k))}$.

In order to prove these lemmas we need to analyze the other graphs as well. More specifically, we will use Lemma 15 to reduce the analysis of the $\mathcal{S}$ graphs to the expansion of the $\mathcal{R}$ and $\mathcal{DS}$ graphs. The bounds we need for Lemma 161 are these.

▷ **Claim 163.** Let $i, j \leq \ell$. Then $\lambda_2(\mathcal{DS}^\ell(i, j)) \leq ij \cdot \gamma^{-(\ell - i - j)}$.

▷ **Claim 164.** Let $U \in \mathcal{M}(\ell)$. The graph $\mathcal{R}^U(i, j)$ is isomorphic to $\mathcal{DS}^\ell(i, \ell - j)$. Consequently by Claim 163 $\lambda_2(\mathcal{R}^\ell(i, j)) \leq \gamma^{-(j - i)}$.

For Lemma 162 we will also need this claim.

▷ **Claim 165.** Let i, j, ℓ be such that $i \leq \frac{j}{2}$ and $n - \ell \geq j$. Let $D \in \mathcal{M}(\ell)$. Then $\lambda_2(\mathcal{R}_{\cap D}(i, j)) \leq \gamma^j$.

The proofs of all these claims are all quite technical, and while most of the proof ideas repeat themselves among claims, we did not find one proof from which to deduce all claims. More details are given in the next subsections. We conclude this sub section with the short proofs of Lemma 47 and Claim 55, that were necessary to prove the main theorem in Section 3. They follow directly from Lemma 161 and Lemma 162.

Proof of Lemma 47. The double cover of $\mathcal{T}^m$ is $\mathcal{S}^{4m}(2m, 2m, m)$. Thus by Lemma 161 and Observation 14 $\lambda(\mathcal{T}^m) = \lambda_2(\mathcal{S}^{4m}(2m, 2m, m)) = \gamma^m$. ◀

Proof of Claim 55. Careful definition chasing reveals that the double cover of $\mathcal{L}(\mathcal{G}(w_1), \mathcal{G}(w_2), W)$ is (isomorphic to) $\mathcal{S}_{\cap Z_W}(2m, 2m, m)$. Indeed, assume for simplicity that $w_1 = w_2 = \mathbb{F}_2^n$. Thus the vertices in the double cover of $\mathcal{L}(\mathcal{G}(w_1), \mathcal{G}(w_2), W)$ are two copies of $\mathcal{M}_{\cap Z_W}(2m)$. This is because by Claim 25, $A \oplus S_W$ if and only if $\text{row}(A) \oplus \text{row}(Z_W), \text{col}(A) \oplus \text{col}(Z_W)$ are a direct sum. Hence by Observation 14 it is enough to argue about the bipartite expansion of $\mathcal{S}_{\cap Z_W}(2m, 2m, m)$.

We observe that $A \sim A'$ in $\mathcal{S}_{\cap Z_W}(2m, 2m, m)$ if and only if $A \sim A'$ in $\mathcal{L}(\mathcal{G}(W_1), \mathcal{G}(W_2), W)$. Matrices A and A' are connected in $\mathcal{L}(\mathcal{G}(w_1), \mathcal{G}(w_2), U)$ when there exists $U \in \mathcal{M}_{\cap Z_W}(4m)$ such that $A \sim A' \in \mathcal{S}^U(2m, 2m, m)$. This is equivalent to existence of $B_1, B_2, B_3, B_4 \in \mathcal{M}(m)$ such that $B_1 \oplus B_2 \oplus B_3 \oplus B_4 \in \mathcal{M}_{\cap Z_W}(4m)$ with $A = B_1 \oplus B_2$ and $A' = B_1 \oplus B_3$ (the sum of B_i 's is U). Observe that B_4 is redundant in this definition, that is, it is enough to require that there exists $B_1, B_2, B_3 \in \mathcal{M}(m)$ such that $B_1 \oplus B_2 \oplus B_3 \in \mathcal{M}_{\cap Z_W}(3m)$ with

$A = B_1 \oplus B_2$ and $A' = B_1 \oplus B_3$. This is because if we are given these B_1, B_2, B_3 as so, we can always find a suitable B_4 . By writing $B_2 = A - B_1$ and $B_3 = A' - B_1$, we get that A, A' are connected in $\mathcal{L}(\mathcal{G}(w_1), \mathcal{G}(w_2), W)$, if and only if $Z_W \oplus B_1 \oplus (A - B_1) \oplus (A' - B_1)$. This is exactly the definition of the edges in $\mathcal{S}_{\cap Z_W}(2m, 2m, m)$.

Thus by Lemma 162 $\lambda(\mathcal{L}(\mathcal{G}(w_1), \mathcal{G}(w_2), W)) = \lambda_2(\mathcal{S}_{\cap Z_W}(2m, 2m, m)) \leq \gamma^{(d_0 - \dim(W) - 3m)} = \gamma^m$. $\triangleleft$

D.1 Expansion of the upper bound local graphs

We begin by showing how Lemma 161 and Lemma 162 reduce to the other claims by Lemma 16.

Proof of Lemma 161. Fix $U \in \mathcal{M}(\ell)$ and i, j, k . We intend to use Lemma 16. We index our components by $\mathcal{M}(k)$, where every component G_C will be the graph containing edges $\{A, B\}$ such that $C \oplus (A - C) \oplus (B - C) \leq U$. We claim that $G_C \cong \mathcal{DS}^{U-C}(i - k, j - k)$, by the map $A \mapsto A - C$. Indeed, by Proposition 28 the vertices in G_C , which are $\mathcal{M}_{\cap C}^U(i) \cup \mathcal{M}_{\cap C}^U(j)$ map isomorphically to the vertices of $\mathcal{DS}^{U-C}(i - k, j - k)$. Moreover, by Claim 154 $C \oplus (A - C) \oplus (B - C) \leq U$ if and only if $(A - C) \oplus (B - C) \leq U - C$.

Thus $G_C \cong \mathcal{DS}^{U-C}(i - k, j - k)$ and by Claim 163 we conclude that $\lambda_2(G_C) \leq \gamma^{(\ell - k - (j - k) - (i - k))} = \gamma^{(\ell - j - (i - k))}$.

The corresponding bipartite decomposition graph of (say) the right-side is the graph $\mathcal{H}_{\tau, R} = (V, T, E)$ given by:

1. $V = \mathcal{M}^U(j)$.
2. $T = \mathcal{M}^U(k)$.
3. $E = \{(A, C) \mid A \geq C\}$.

This is precisely $\mathcal{R}^U(j, k)$. By Claim 164 this graph is an $\gamma^{(\ell - j - k)}$ -spectral expander. Therefore by Lemma 16, $\lambda_2(\mathcal{S}^U(i, j, k)) \leq \gamma^{(\ell - j - \max(k, i - k))}$. $\triangleleft$

Proof of Lemma 162. We use Lemma 16. Our components will be $G^U = \mathcal{S}^U(i, j, k)$ for all $U \in \mathcal{M}_{\cap D}(n - \ell)$. We observe that sampling an edge by sampling $U \in \mathcal{M}_{\cap D}(n - \ell)$ and then sampling an edge in $\mathcal{S}^U(i, j, k)$, results in a uniform distribution over the edges of $\mathcal{M}_{\cap D}(i, j, k)$. By Lemma 161 every such component has $\lambda_2(G^U) \leq \gamma^{(n - \ell - j - \max(k, i - k))}$.

We note that the (say) left side of the bipartite decomposition graph is the bipartite graph $\mathcal{H}_{\tau, L} = (V, T, E)$ with:

1. $V = \mathcal{M}_{\cap D}(i)$.
2. $T = \mathcal{M}_{\cap D}(n - \ell)$.
3. $E = \{(B, U) \mid B \leq U\}$.

This is precisely $\mathcal{R}_{\cap D}(i, n - \ell)$. By Claim 165, $\lambda_2(\mathcal{R}_{\cap D}(i, n - \ell)) \leq \gamma^{(n - \ell - i)} \leq \gamma^{(n - \ell - j - \max(k, i - k))}$ so by Lemma 16,

$$\lambda_2(\mathcal{S}_{\cap D}(i, j, k)) \leq \gamma^{(n - \ell - j - \max(k, i - k))}. \quad \triangleleft$$

Let us also give a short proof for Claim 164.

Proof of Claim 164. Fix i, j, ℓ and $U \in \mathcal{M}(\ell)$. For notational convenience let $\mathcal{R}^U(i, j) = (L, R, E)$ and $\mathcal{DS}^U(i, \ell - j) = (L', R', E')$. We define $\phi : L \cup R \rightarrow L' \cup R'$ to be

$$\forall B \in L, \phi(B) = B \quad \text{and} \quad \forall A \in R, \phi(A) = U - A.$$

Let us first show that indeed ϕ bijectively maps the vertices L, R to L', R' respectively. For L this is clear. Moreover, $A \leq U$ if and only if $U - A \leq U$, and in this case $\text{rank}(A) = j$ if and only if $\text{rank}(U - A) = \text{rank}(U) - \text{rank}(A) = \ell - j$. Thus ϕ is a bijection of the vertex sets in every side.

Next we verify that $\{A, B\} \in E$ if and only if $\{U - A, B\} \in E'$. Indeed, $\{A, B\} \in E$ if and only if $B \leq A \leq U$. By definition of $\oplus$, this is if and only if $A = B \oplus (A - B)$ and $U = A \oplus (U - A)$, or more compactly, $B \oplus (A - B) \oplus (U - A) = U$. On the other hand, $B \oplus (U - A) \leq U$ if and only if $B \oplus (U - A) \oplus (A - B) = U$, and thus $\{A, B\} \in E$ if and only if $\{U - A, B\} \in E'$. The isomorphism is proven. $\triangleleft$

In the next subsections we give proofs to Claim 163 and Claim 165, are more technical.

D.2 Bounding the expansion of the direct sum graphs

The proof of Claim 163 is by induction on the sum of matrix ranks, $i + j$. We begin with the base case.

▷ Claim 166. $\lambda_2(\mathcal{DS}^U(1, 1)) \leq O(2^{-\frac{1}{2}\ell})$.

The base case follows from the fact that the walk in $\mathcal{DS}^U(1, 1)$ is close to the uniform random walk as in Corollary 11.

Proof of Claim 166. Without loss of generality let us assume that $U = I$, the identity matrix whose rank is ℓ . For vectors $x, y \in \mathbb{F}_2^\ell$ $\langle x, y \rangle = \sum_{j=1}^\ell x_j y_j$ the standard bilinear form over $\mathbb{F}_2$, which we sometimes abuse notation and call an “inner product”. Golowich observes the following

▷ Claim 167 ([50, Lemma 66]). Let I the identity matrix of rank ℓ . A matrix $A \leq I$ if and only if $A = \sum_{i=1}^{\ell''} e_i \otimes f_i$ for some $\ell'' \leq \ell$ and $\{e_i\}_{i=1}^{\ell''}, \{f_i\}_{i=1}^{\ell''}$ such that

$$\langle e_i, f_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Thus $\mathcal{DS}^U(1, 1)$ is (isomorphic to) the bipartite graph where $L = \{(e, f) \in \mathbb{F}_2^{2\ell} \mid \langle e, f \rangle = 1\}$, R is a disjoint copy of L , and we connect $(e, f) \sim (e', f')$ if and only if $\langle e, f' \rangle = \langle e', f \rangle = 0$ (i.e. we move from $e \otimes f$ to (e, f)). We will use Corollary 11 and show that the two step walk in this graph is $O(2^{-\ell})$ -close to the uniform graph, thus proving that $\lambda_2(\mathcal{DS}^U(1, 1)) \leq O(2^{-\frac{1}{2}\ell})$. For concreteness we will use the two step walk from L to L .

Let us fix an initial vertex (e_0, f_0) and let μ be the distribution over vertices in the two step walk. Let J be the uniform distribution over vertices. We want to show that $\|\mu - J\| = \sum_{v \in V} |\mu(v) - \frac{1}{|V|}| \leq O(2^{-\ell})$.

The basic idea is this: we prove below that the probability that we traverse from (e_0, f_0) to (e_2, f_2) essentially only depends on whether $e_0 = e_2$, and whether $f_0 = f_2$ (and not on e_2 or f_2 themselves). While this is not true *per se*, it is true up to some negligible probability. Then we use the large regularity to bound the probability that (e_2, f_2) has $e_0 = e_2$ or $f_0 = f_2$ by $O(2^{-\ell})$. This asserts that μ is (almost) uniform on a set that contains all but a negligible fraction of the vertices, which is enough to show closeness to J . Details follow.

We record for later the following observations.

► Observation 168.

1. If $(e, f) \in L$ then $e, f \neq 0$.
2. $|L| = (2^\ell - 1)2^{\ell-1}$.
3. The graph is $D = (2^{\ell-1} - 1)2^{\ell-2}$ -regular.

The first two observations are immediate, but we explain the third: fixing (e, f) , we count its neighbors (e', f') . There are $2^{\ell-1} - 1$ choices for e' since it is perpendicular to f and non-zero. Given e' , the vector f' satisfies

$$\langle e', f' \rangle = 1 \quad \text{and} \quad \langle e, f' \rangle = 0.$$

Note that e, e' are linearly independent since both are non-zero and $\langle e, f \rangle \neq \langle e', f \rangle$. There are $2^{\ell-2}$ solutions to these linear equations hence the graph is $(2^{\ell-1} - 1)2^{\ell-2}$ -regular.

Let (e_2, f_2) be any vertex. The first crude bound we give on the is just by regularity. For any $(x, y) \in V$, $\mu(x, y) \leq \frac{1}{D} = \frac{1}{(2^{\ell-1}-1)2^{\ell-2}}$. This bound is tight when $(e_0, f_0) = (e_2, f_2)$. Similarly, if either $e_0 = e_2$ or $f_0 = f_2$ then we also just bound $\mu(e_0, f_0) \leq \frac{1}{(2^{\ell-1}-1)2^{\ell-2}}$; there are at most 2^ℓ such pairs.

Now let us assume $e_2 \neq e_0, f_2 \neq f_0$ and calculate more accurate upper and lower bounds on $\mu(e_2, f_2)$ by counting how many two paths there are from (e_0, f_0) to (e_2, f_2) , or equivalently, how many $(e_1, f_1) \in R$ are common neighbors of $(e_0, f_0), (e_2, f_2)$.

Given e_0, e_2 , the vector e_1 needs to satisfy $\langle f_0, e_1 \rangle = \langle f_2, e_1 \rangle = 0$. The two vectors f_0, f_2 are linearly independent by assumption that they are both distinct and non-zero. Thus there are $2^{\ell-2} - 1$ non-zero e_1 -s that satisfy $\langle f_0, e_1 \rangle = \langle f_2, e_1 \rangle = 0$.

Fixing e_1 the vector f_1 needs to satisfy $\langle e_0, f_1 \rangle = \langle e_2, f_1 \rangle = 0$ and $\langle e_1, f_1 \rangle = 1$. If e_0, e_1, e_2 are linearly independent then there are exactly 2^{m-3} solutions to these equations, all of them are non-zero, and every such solution corresponds to a common neighbor (e_1, f_1) . If they are not linearly independent we claim that there is no solution. We note that every pair must be linearly independent: All three are non-zero vectors. Moreover, e_0, e_2 are distinct vectors by assumption, and similarly $e_1 \neq e_0$ (resp. $e_1 \neq e_2$) because $\langle f_0, e_0 \rangle \neq \langle f_0, e_1 \rangle$ (resp. $\langle f_2, e_2 \rangle \neq \langle f_2, e_1 \rangle$). Every pair of distinct non-zero vectors are independent. Thus, if there is a dependency then $e_1 = e_0 + e_2$, and so the first two equations imply that $\langle e_1, f_1 \rangle = 0$, so there is no solution.

We conclude that the number of paths from (e_0, f_0) to (e_2, f_2) is one of the following.

1. $(2^{\ell-2} - 1)2^{\ell-3} = 2^{2\ell-5} - 2^{\ell-3}$ - if $e_0 + e_2$ is not a valid solution in the first step, i.e. either $\langle f_0, e_0 + e_2 \rangle \neq 0$ or $\langle f_2, e_0 + e_2 \rangle \neq 0$.
2. $(2^{\ell-2} - 2)2^{\ell-3} = 2^{2\ell-5} - 2^{\ell-2}$ otherwise.

We have that for every such pair,

$$\mu(e_2, f_2) = \frac{2^{2\ell-5}}{((2^{\ell-1}-1)2^{\ell-2})^2} - \frac{2^{\ell-b}}{((2^{\ell-1}-1)2^{\ell-2})^2},$$

for $b \in \{2, 3\}$. This implies that $\mu(e_2, f_2) = \frac{1}{|V|} + \varepsilon_1 + \varepsilon_2$ for

$$|\varepsilon_2| = \left| \frac{2^{\ell-b}}{((2^{\ell-1}-1)2^{\ell-2})^2} \right| \leq 2^{6-3\ell}$$

and

$$\begin{aligned} |\varepsilon_1| &\leq \left| \frac{2^{2\ell-4}}{((2^{\ell-1}-1)2^{\ell-2})^2} - \frac{1}{|V|} \right| \\ &= \frac{2(3 \cdot 2^\ell - 5)}{(2^\ell - 2)^2(2^\ell - 1)2^\ell} \leq 2^{6-3\ell} \end{aligned}$$

Thus in particular,

$$\|\mu - J_{(e_0, e_2)}\|_1 = \sum_{e_2=e_0 \vee f_2=f_0} |\mu(e_2, f_2) - \frac{1}{|V|}| + \sum_{e_2 \neq e_0 \text{ and } f_2 \neq f_0} |\mu(e_2, f_2) - \frac{1}{|V|}|$$

$$\begin{aligned}
&\leq \frac{2^\ell}{(2^{\ell-1} - 1)2^{\ell-2}} + 2^{2\ell}(\varepsilon_1 + \varepsilon_2) \\
&\leq 2^{-\ell}(2^4 + 2^7) \leq 2^8 2^{-\ell}.
\end{aligned}$$

The claim now follows from Corollary 11. $\triangleleft$

Now we prove Claim 163. As mentioned before, the proof is by induction on $i + j$. A similar induction was done in [26] when analyzing the so-called *swap walk*, a generalization of the Kneser graph to high dimensional expanders. In the inductive step, one uses Lemma 16 to reduce the analysis of (i, j) to that of $(i - 1, j)$ (which holds true by the inductive assignment). Details follow.

Proof of Claim 163. We show that if $i > 1$ then

$$\lambda_2(\mathcal{DS}^\ell(i, j)) \leq \lambda_2(\mathcal{DS}^{\ell-1}(i - 1, j)) + \lambda_2(\mathcal{DS}^\ell(1, j)). \quad (\text{D.1})$$

By using this inequality we eventually get that

$$\lambda_2(\mathcal{DS}^\ell(i, j)) \leq (i \cdot j) \lambda_2(\mathcal{DS}^{\ell-i-j}(1, 1)) = O\left((i \cdot j) 2^{-\frac{1}{2}(\ell-i-j)}\right),$$

where the final equality holds by Claim 166.

To prove (D.1) we use Lemma 16. In this case our local decomposition is $T = \{G_B \mid B \in \mathcal{M}^U(1)\}$. The graph $G_B = (L_B, R_B)$ is the induced bipartite subgraph where $L_B = \{A \in L \mid B \leq A \leq U\}$ and $R_B = \{A \in R \mid (A, B) \in \mathcal{DS}^U(1, j)\}$. We claim that $G_B \cong \mathcal{DS}^{U-B}(i - 1, j)$: the isomorphism $\phi : G_B \rightarrow \mathcal{DS}^{U-B}(i - 1, j)$ maps

$$A \in R_B \mapsto \phi(A) = A \quad \text{and} \quad A \in L_B \mapsto \phi(A) = A - B.$$

For the right sides, we note that $(A, B) \in \mathcal{DS}^U(1, j)$ if and only if $A \leq U - B$ so this isomorphism maps the right side of G_B to that of $\mathcal{DS}^{U-B}(i - 1, j)$. For the left sides, Proposition 28 gives us that $B \leq A \leq U$ if and only if $A - B \leq U - B$. Thus ϕ maps matrices $A \in R_B$ isomorphically to matrices on the left side of $G_{m-1}^0(U - B, a_1, a_2 - 1)$. Checking that $A_1 \sim A_2$ in $\mathcal{DS}^{U-B}(i - 1, j)$ if and only if $\phi(A_1) \sim \phi(A_2)$ in G_B is direct. Therefore for every B , the second largest eigenvalue $\lambda_2(G_B) = \lambda_2(\mathcal{DS}^{\ell-1}(i - 1, j))$.

We note that by definition the graph $\mathcal{H}_{\tau, L} = \mathcal{DS}^\ell(1, j)$. Thus (D.1) follows from Lemma 16. $\triangleleft$

D.3 Bounding the expansion of the relation graphs

We start by proving the claim in the following special case.

$\triangleright$ Claim 169. Let $D \in \mathcal{M}(\ell)$ and let i, j be such that $i \leq \frac{j}{3}$, then $\lambda_2(\mathcal{R}_{\cap D}(i, j)) \leq \gamma^j$.

After proving Claim 169, we prove Claim 165 by using Lemma 16 again.

Proof of Claim 169. Fix D, i, j . For this case we intend to use Corollary 11 on the two step walk on $\mathcal{M}_{\cap D}(i)$. We prove that

$$\max_{B_1 \in V} \|\mu_{B_1} - J\|_1 \leq \gamma''^j \quad (\text{D.2})$$

where μ_{B_1} is the distribution of the two step walk starting from B_1 , and the distribution J is the uniform distribution over $\mathcal{M}_{\cap D}(i)$. Fix $B_1 \in \mathcal{M}_{\cap D}(i)$ and from now on we just write μ instead of μ_{B_1} for short.

We will show that there exists a set $V' \subseteq \mathcal{M}_{\cap D}(i)$ such that both $\mathbb{P}_{B_2 \sim \mu}[B_2 \in V'], \mathbb{P}_{B_2 \sim J}[B_2 \in V'] \geq 1 - p$ for a sufficiently small $p \leq (\gamma')^j$. Then we will also prove that the distribution μ conditioned on this event V' is uniform, and this will give us (D.2). Here

$$V' = \{B_2 \in \mathcal{M}_{\cap D}(i) \mid \text{row}(B_2) \cap (\text{row}(B_1) \oplus \text{row}(D)) = \{0\}, \\ \text{col}(B_2) \cap (\text{col}(B_1) \oplus \text{col}(D)) = \{0\}\}.$$

A quick calculation will show that in this case $\|\mu - J\|_1 \leq 2p \leq \gamma^j$ for some $\gamma \in (0, 1)$.

First we show that $\mathbb{P}_{B_2 \sim J}[V'] \geq 1 - O(2^{2i-j})$: Let us show that the probability of choosing a uniform $B_2 \in \mathcal{M}_{\cap D}(i)$ such that $\text{row}(B_2) \cap (\text{row}(B_1) \oplus \text{row}(D)) \neq \{0\}$ is at most $i2^{2i-j}$ (the same also holds for the column spaces). The row span of B_2 is sampled uniformly as an i dimensional subspace inside a n dimensional space, conditioned on $\text{row}(B_2) \cap \text{row}(D) = \{0\}$. So this probability is

$$\mathbb{P}_{\text{row}(B_2)}[(\text{row}(B_1) \oplus \text{row}(D)) \cap \text{row}(B_2) \neq \{0\} \mid \text{row}(B_2) \cap \text{row}(D) = \{0\}] \\ \leq \frac{\mathbb{P}_{\text{row}(B_2)}[(\text{row}(B_1) \oplus \text{row}(D)) \cap \text{row}(B_2) \neq \{0\}]}{\mathbb{P}_{\text{row}(B_2)}[\text{row}(B_2) \cap \text{row}(D) = \{0\}]}$$

The ratio is bounded by $\frac{i2^{2i+\ell-n-1}}{1-i2^{\ell-n}} \leq O(i2^{2i+(\ell-n)})$ via standard calculations. The same probability holds for the columns spaces, therefore $\mathbb{P}_{B_2 \sim J}[V'] \geq 1 - O(i2^{i+(\ell-n)}) = 1 - O(\gamma^{(2i-j)})$. The last equality is by $j \leq n - \ell$.

Next we claim that $\mathbb{P}_{B_2 \sim \mu}[V'] \geq 1 - O(2^{2i-j})$. Indeed, the two step walk from B_1 traverses to some $U \in \mathcal{M}_{\cap D}(j)$ (whose row space intersects trivially with $\text{row}(D)$), and from U we independently choose a matrix $B_2 \leq U$. As $\text{row}(U) \cap \text{row}(D) = \{0\}$, it follows that $\text{row}(B_2) \cap \text{row}(B_1) \oplus \text{row}(D) \neq \{0\}$ if and only if $\text{row}(B_1) \cap \text{row}(B_2) \neq \{0\}$. For any fixed $U \in \mathcal{M}_{\cap D}(j)$, the probability that $\text{row}(B_1) \cap \text{row}(B_2) \neq \{0\}$ is at most $O(i2^{2i-j})$ (repeating the calculation above).

Now let us show that conditioned on V' , for any two neighbors $B_2, B'_2 \in V'$ of B_1 in the two step walk, the probability of sampling B_2 is the same as sampling B'_2 . To show this it is enough to show that there exists a graph isomorphism of the original “one-step graph”, that fixes B_1 and sends B_2 to B'_2 . Let us denote by $GL(\mathbb{F}_2^n)$ the general linear group, or equivalently the group of $n \times n$ invertible matrices. We will find $E_1 \in GL(w_1), E_2 \in GL(w_2)$ such that:

1. $E_2 A = A E_1 = A$.
2. $E_2 B_1 E_1 = B_1$.
3. $E_2 B_2 E_1 = B'_2$.

Assume we found such E_1, E_2 . We claim that $X \mapsto E_2 X E_1$ is an isomorphism of the one-step graph. Indeed, note that if $A \oplus B$ then as multiplying by an invertible matrix maintains rank, it follows that $E_2 A E_1 \oplus E_2 B E_1 = A \oplus B E_1$. Therefore, $X \mapsto E_2 X E_1$ is a bijection on the vertices of the graph. It is easy to check that this multiplication also maintains the $\leq$ relation so this is indeed an isomorphism. Finally, it is also easy to see that (B_1, M, B_2) is mapped to a path $(B_1, E_2 M E_1, B'_2)$, so in particular such an isomorphism implies that the probability of sampling $B_2, B'_2 \in V'$ is the same.

To find such E_1, E_2 recall that $E_2(e \otimes f)E_1 = (E_2 e) \otimes (E_1 f)$. Thus we can decompose $A = \sum_{t=1}^{\ell} e_t \otimes f_t$, $B_1 = \sum_{t=1}^i e'_t \otimes f'_t$, $B_2 = \sum_{t=1}^i h_t \otimes k_t$ and $B'_2 = \sum_{t=1}^i h'_t \otimes k'_t$. By assumption on the intersections of the row (and column) spaces, the sets $\{e_t\}_{t=1}^{\ell} \cup \{e'_t\}_{t=1}^i \cup \{h_t\}_{t=1}^i$ and

$\{e_t\}_{t=1}^\ell \cup \{e'_t\}_{t=1}^i \cup \{h'_t\}_{t=1}^i$ are independent and have the same size. The same holds for the corresponding sets of column vectors (replacing e, h with f, k). Thus we just find any E_1, E_2 such that

$$E_1 e_t = e_t, E_1 e'_t = e'_t, E_1 h_t = h'_t$$

$$E_2 f_t = f_t, E_1 f'_t = f'_t, E_1 k_t = k'_t.$$

These will have the desired properties. By the arguments above, the claim is proven. $\triangleleft$

We comment that the above proof is one place where an analogous argument fails to hold for higher dimensional tensors. We heavily relied on the fact that for any fixed matrix U , the row span and column span of an i rank matrix $B \leq U$ is uniformly distributed over i dimensional subspaces inside $\text{row}(U)$ and $\text{col}(U)$, even if these marginals are not independent.

We move on to the general case.

Proof of Claim 165. Fix i, j, ℓ and $D \in \mathcal{M}(\ell)$. The case where $i \leq \frac{j}{3}$ is by Claim 169 so assume that $\frac{j}{3} < i \leq \frac{j}{2}$. Towards using Lemma 16, we decompose $\mathcal{R}_{\cap D}(i, j)$ to components $\{G_B\}_{B \in \mathcal{M}_{\cap D}(\frac{j}{3})}$ where G_B is the graph induced by all edges $\{M, C\}$ such that $B \leq C \leq M$. We claim that G_B is isomorphic to $\mathcal{R}_{\cap(D \oplus B)}(i - \frac{j}{3}, \frac{2j}{3})$. The isomorphism $\phi : G_B \rightarrow \mathcal{R}_{\cap(D \oplus B)}(i - \frac{j}{3}, \frac{2j}{3})$ is $\phi(M) = M - B$. Let us start by showing that this maps vertices to vertices.

Indeed, fix any $B \in \mathcal{M}_{\cap D}(\frac{j}{3})$ and let $G_B = (L_B, R_B, E_B)$. Explicitly,

$$L_B = \{M \in \mathcal{M}_{\cap D}(i) \mid M \geq B\}, R_B = \{C \in \mathcal{M}_{\cap D}(j) \mid C \geq B\}.$$

Note that every $M \in L_B$ is a direct sum $M \oplus D$, and as $B \leq M$ this implies that $(B \oplus (M - B)) \oplus D$. As we saw before, this implies that $(B \oplus D) \oplus (M - B)$, therefore $\phi(M) = M - B \in \mathcal{M}_{\cap(D \oplus B)}(i - \frac{j}{3})$. Conversely, if $M' \in \mathcal{M}_{\cap(D \oplus B)}(i - \frac{j}{3})$ then $\phi^{-1}(M') = M' \oplus B$ has $\text{rank}(\phi^{-1}(M')) = i$. Moreover, as $M' \oplus (D \oplus B)$ then $(M' \oplus B) \oplus D$, i.e. $\phi^{-1}(M') \in \mathcal{M}_{\cap D}(j)$. It is also clear that $B \leq \phi^{-1}(M')$ hence $\phi^{-1}(M') \in L_B$. Of course, a similar argument holds for the right side R_B , then ϕ is a bijection of the vertex set. Finally, by Claim 154, $C \leq M$ if and only if $C - B \leq M - B$, so this is indeed an isomorphism of graphs.

We note that if $i \leq \frac{j}{2}$ then $i - \frac{j}{3} \leq \frac{1}{2} \cdot \frac{2j}{3}$ and therefore we can use Claim 169 and deduce that for every $B \in \mathcal{M}_{\cap D}(\frac{j}{3})$, $\lambda_2(G_B) \leq \gamma^j$. Moreover, we observe that the bipartite local to global graph for the left side, $\mathcal{H}_{\tau, L}$ is $\mathcal{M}_{\cap D}(\frac{j}{3}, j)$, thus by Claim 169 this also has $\lambda_2(\mathcal{H}_{\tau, L}) \leq \gamma^j$ and so by Lemma 16 the lemma is proven. $\triangleleft$

We comment that the requirement that $i \leq \frac{j}{2}$ is just for convenience. Multiple uses of the same decomposition with Lemma 16, would yield a similar bound whenever $i \leq (1 - \varepsilon)j$ for any fixed constant $\varepsilon > 0$. As we only need the claim for $i = \frac{j}{2}$, we merely mention this without proof.

D.4 Short paths in $\mathcal{T}^m$

This subsection of the appendix is devoted to the proof of Claim 117 which is necessary to analyze the coboundary expansion of the complexes in [50].

We begin with the following auxiliary claim proven in the end of the subsection.

$\triangleright$ Claim 170. Let Z_1 be a rank $4m$ matrix and let Z_2 be a rank $3m$ matrix.

1. Let $A, B \leq Z_1$ be such that $\text{rank}(A) = \text{rank}(B) = m$. Then there exists a matrix T with $\text{rank}(T) = m$ such that $A \oplus T, T \oplus B \leq Z_1$.

2. Let $A, B \leq Z_2$ be such that $\text{rank}(A) = \text{rank}(B) = m$ then there exists M^1, M^2, M^3 such that

$$A \oplus M^1, M^1 \oplus M^2, M^2 \oplus M^3, M^3 \oplus B \leq Z_2.$$

Phrasing this claim differently, the first item claims that any two matrices in one side of $\mathcal{DS}^{4m}(m, m)$ have a length 2 path between them. The first item claims that any two matrices in one side of $\mathcal{DS}^{3m}(m, m)$ have a length 4 path between them.

Proof of Claim 117. We proceed as follows. We first decompose $C, D \in \mathcal{T}^Z$ into m ranked matrices such that $C = C^1 \oplus C^2$ and $D = D^1 \oplus D^2$. Then we find some T of rank m such that $T \oplus C^1, T \oplus D^1 \leq Z$. This is possible due to the first item in Claim 170.

Then we will prove that there exists a 12-path composed of three paths of length 4: one from $C^1 \oplus C^2$ to $C^1 \oplus T$, one from $C^1 \oplus T$ to $D^1 \oplus T$ and one from $D^1 \oplus T$ to $D^1 \oplus D^2$.

Let us show how to construct the first path, the other two are constructed similarly. Note that $C^1 \leq C^1 \oplus C^2, C^1 \oplus T$ and that they are both $\leq Z$. Thus if we find $M^1, M^2, M^3 \leq Z - C^1$ such that

$$C^2 \oplus M^1, M^1 \oplus M^2, M^2 \oplus M^3, M^3 \oplus T \leq Z - C^1,$$

Then by Proposition 28, adding C^1 to all three M^1, M^2, M^3 will give us the desired path in $\mathcal{T}^Z$:

$$(C^1 \oplus C^2, C^1 \oplus M^1, C^1 \oplus M^2, C^1 \oplus M^3, C^1 \oplus T).$$

But observe that $Z - C^1$ has rank $3m$, and thus existence of such matrices M^1, M^2, M^3 is promised by Claim 170 so we can conclude. $\triangleleft$

Proof of Claim 170. Without loss of generality let us assume that Z_1 is the identity matrix of rank $4m$ and that Z_2 is the identity matrix of rank $3m$. We wish to use the criterion for $A \leq I$ given in Claim 167.

For this proof we also require the following observation which follows from elementary linear algebra. We leave its proof to the reader.

► **Observation 171.** Let $W \subseteq \mathbb{F}_2^n$ be a subspace of dimension j . Let $E = \{e_1, e_2, \dots, e_k\}$ be k independent vectors. Assume that $W \cap \text{span}(E) = \{0\}$. Then for every $b_1, b_2, \dots, b_k \in \mathbb{F}_2$ there exists a solution x to the set of linear equations:

1. For every $w \in W$, $\langle x, w \rangle = 0$, and
2. For every $i = 1, 2, \dots, k$, $\langle x, e_i \rangle = b_i$.

Let us begin with the first item. Given A, B of rank m we need to find some T of rank m such that $A \oplus T, B \oplus T \leq Z_1$. Equivalently, we will find T such that $A \oplus T$ and $B \oplus T$ satisfies the criterion in Claim 167.

Let us write $A = \sum_{j=1}^m e_j(A) \otimes f_j(A), B = \sum_{j=1}^m e_j(B) \otimes f_j(B)$ where these decompositions satisfy the property in Claim 167. Let $W = \{w \in \mathbb{F}_2^{4m} \mid \forall j = 1, 2, \dots, m \langle w, f_j(A) \rangle = \langle w, f_j(B) \rangle = 0\}$. As W is a solution space of $2m$ linear equations over $\mathbb{F}_2^{4m}$, it is a subspace of dimension $\geq 2m$. We will select the rows of T from W .

However, it is not enough to just select arbitrary rows from W , since we need to satisfy the condition that if e_i, f_j participates in the decomposition of T then $\langle e_i, f_j \rangle = 1$ if and only if $i = j$. In addition, we will also require that f_j will be perpendicular to all the $e_i(A), e_i(B)$'s.

This will define a set of linear equations which we will solve using Observation 171. To do this we will find an m -dimensional subspace in $W' \subseteq W$ that intersects trivially with the span of the $e_j(A), e_j(B)$'s (the sum of row spaces of A and B).

We note that $W \cap \text{row}(A) = W \cap \text{row}(B) = \{0\}$. Let us show this for $\text{row}(A)$: any non-zero vector $x \in \text{row}(A)$ is a sum of $\sum_{j=1}^m b_j e_j(A)$ where at least one of the b_j 's are not zero. On the other hand, $\langle x, f_j(A) \rangle = b_j$, so there is at least one $\langle x, f_j(A) \rangle \neq 0$, implying that $x \notin W$. Thus we conclude that $\dim(W \cap (\text{row}(A) + \text{row}(B))) \leq m$, and in particular there exists some m dimensional subspace $W' \subseteq W$ such that W' intersects trivially with $\text{row}(A) + \text{row}(B)$. Let $E = \{e_1(T), \dots, e_m(T)\}$ be a basis of W' .

Thus by Observation 171, for every $i = 1, 2, \dots, m$ there exists some $f_i(T)$ such that

1. For every $j = 1, 2, \dots, m$, $\langle e_j(A), f_i(T) \rangle = \langle e_j(B), f_i(T) \rangle = 0$ for all $j = 1, 2, \dots, m$, and
2. In addition

$$\langle e_j(T), f_i(T) \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

By taking $T = \sum_{j=1}^m e_j(T) \otimes f_j(T)$ we observe that both $A \oplus T$ and $B \oplus T$ satisfy Claim 167, and therefore $A \oplus T, B \oplus T \leq Z_1$.

Let us turn to the second item. We continue with the notation $A = \sum_{j=1}^m e_j(A) \otimes f_j(A), B = \sum_{j=1}^m e_j(B) \otimes f_j(B)$. The strategy here will be similar. Let $W^1 = \{w \in \mathbb{F}_2^{3m} \mid \forall j = 1, 2, \dots, m, \langle w, f_j(A) \rangle = \langle w, f_j(B) \rangle = 0\}$. We observe that this is the solution set of $2m$ linear equations and therefore its dimension is $\geq m$. Let $e_1(M^1), e_2(M^1), \dots, e_m(M^1) \in W^1$ be m independent vectors, and set $e_j(M^3) = e_j(M^1)$. We observe that as before $W^1 \cap \text{row}(A) = \{0\}$ so by Observation 171, for every $j = 1, 2, \dots, m$ there exists some $f_j(T)$ such that

1. For every $e_i(A)$, $\langle e_i(A), f_j(M^1) \rangle = 0$.
2. For every i, j ,

$$\langle e_i(M^1), f_j(M^1) \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

We note that the linear equations for $f_j(M^1)$'s do not consider the $e_j(B)$'s.

Similarly, we find $f_j(M^3)$ ignoring the $e_j(A)$'s. That is, for every $j = 1, 2, \dots, m$ we find $f_j(M^3)$ such that

1. For every $e_i(B)$, $\langle e_i(B), f_j(M^3) \rangle = 0$.
2. For every i, j ,

$$\langle e_i(M^3), f_j(M^3) \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Then we set $M^1 = \sum_{j=1}^m e_j(M^1) \otimes f_j(M^1)$ and $M^3 = \sum_{j=1}^m e_j(M^3) \otimes f_j(M^3)$. It is easy to verify that as before $A \oplus M^1, B \oplus M^3 \leq Z_2$.

Now we find M^2 such that $M^1 \oplus M^2, M^2 \oplus M^3 \leq Z_2$. The upshot of the previous step is that now M^1 and M^3 have the same row space $\text{span}(\{e_1(M^1), e_2(M^1), \dots, e_m(M^1)\})$. Thus we find M^2 as follows. Let $W^2 = \{w \in \mathbb{F}_2^{3m} \mid \forall j = 1, 2, \dots, m, \langle w, f_j(A) \rangle = \langle w, f_j(B) \rangle = 0\}$. This subspace also has dimension $\geq m$ so let $e_1(M^2), e_2(M^2), \dots, e_m(M^2) \in W^2$ be a set of independent vectors from W^2 .

As before, we observe that $W^2 \cap \text{row}(M^1) = \{0\}$ but this time $\text{row}(M^1) + \text{row}(M^3) = \text{row}(M^1)$. Thus by Observation 171 for every $j = 1, 2, \dots, m$, we can find $f_j(M^2)$ such that

1. For every $i = 1, 2, \dots, m$, $\langle e_i(M^1), f_j(M^2) \rangle = \langle e_i(M^3), f_j(M^2) \rangle = 0$, and
2. In addition,

$$\langle e_i(M^2), f_j(M^2) \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

Then we set $M^2 = \sum_{j=1}^m e_j(M^2) \otimes f_j(M^2)$. It is direct to verify that $M^1 \oplus M^2, M^2 \oplus M^3 \leq Z_2$ by Claim 167. $\triangleleft$

E Second largest eigenvalue of $J(n, k, k/2)$

Proof of Lemma 65. Let $n = (2 + \eta)k$. For the Johnson graph $J(n, k, k/2)$, its largest unnormalized eigenvalue $\lambda_0 = \binom{k}{k/2} \binom{n-k}{k/2}$. Therefore by Theorem 63 the absolute values of the normalized eigenvalues are

$$\begin{aligned} \frac{|\lambda_t|}{\lambda_0} &= \frac{|\sum_{i=0}^t (-1)^{t-i} \binom{k-i}{k/2-i} \binom{n-k+i-t}{k/2+i-t} \binom{t}{i}|}{\binom{k}{k/2} \binom{n-k}{k/2}} \\ &= \left| \sum_{i=\max(0, t-k/2)}^{\min(t, k/2)} (-1)^{t-i} \binom{t}{i} \cdot \frac{k/2 \times \dots \times (k/2 - i + 1)}{k \times \dots \times (k - i + 1)} \right. \\ &\quad \left. \cdot \frac{k/2 \times \dots \times (k/2 + i - t + 1)}{(n-k) \times \dots \times (n-k + i - t + 1)} \right|. \end{aligned}$$

Here the convention is that when we write $\binom{a}{b}$ for $b > a$ or $b < 0$, then the term is equal to 0. Thus, the terms corresponding to $i < t - k/2$ or $i > \min\{t, k/2\}$ vanish since the product of the three binomial coefficients is 0 for these choices of i . We first show the statement for odd $t \in [k]$. In this case, we group terms $\binom{t}{i}$ and $\binom{t}{t-i}$ together to get

$$\begin{aligned} &\leq \sum_{i=\max(0, t-k/2)}^{\lfloor t/2 \rfloor} \binom{t}{i} \cdot \left| \prod_{j=0}^{i-1} \frac{(k/2-j)}{(k-j)} \cdot \prod_{j=0}^{t-i-1} \frac{k/2-j}{n-k-j} - \prod_{j=0}^{t-i-1} \frac{k/2-j}{k-j} \cdot \prod_{j=0}^{i-1} \frac{k/2-j}{n-k-j} \right| \\ &= \sum_{i=\max(0, t-k/2)}^{\lfloor t/2 \rfloor} \binom{t}{i} \cdot \prod_{j=0}^{i-1} \frac{k/2-j}{k-j} \cdot \prod_{j=0}^{i-1} \frac{k/2-j}{n-k-j} \cdot \left| \prod_{j=i}^{t-i-1} \frac{k/2-j}{n-k-j} - \prod_{j=i}^{t-i-1} \frac{k/2-j}{k-j} \right|. \end{aligned}$$

To simplify the products we first observe that for any $0 \leq j < k/2$, $\frac{k/2-j}{k-j} \leq \frac{k/2}{k} = \frac{1}{2}$ and $\frac{k/2-j}{n-k-j} \leq \frac{k/2}{n-k} = \frac{1}{2(1+\eta)}$.

$$\leq \sum_{i=\max(0, t-k/2)}^{\lfloor t/2 \rfloor} \binom{t}{i} \cdot \left(\frac{1}{2}\right)^i \cdot \left(\frac{1}{2(1+\eta)}\right)^i \cdot \left| \prod_{j=i}^{t-i-1} \frac{k/2-j}{n-k-j} - \prod_{j=i}^{t-i-1} \frac{k/2-j}{k-j} \right|. \quad (\text{E.1})$$

Furthermore we use the following claim to bound the term in absolute value:

$\triangleright$ **Claim 172.** For any constant $\eta > 0$ and any integers $1 \leq m < C$ and $0 \leq s \leq m$,

$$\prod_{j=s}^m \frac{C-j}{2C-j} - \prod_{j=s}^m \frac{C-j}{2(1+\eta)C-j} \leq \left(\frac{C-s}{2C-s}\right)^{m-s+1} - \left(\frac{C-s}{2(1+\eta)C-s}\right)^{m-s+1}.$$

A proof is given later in the appendix. By the claim the expression in Equation (E.1) can be bounded by a binomial sum:

$$\begin{aligned}
&\leq \sum_{i=\max(0, t-k/2)}^{\lfloor t/2 \rfloor} \binom{t}{i} \cdot \left(\frac{1}{2}\right)^i \cdot \left(\frac{1}{2(1+\eta)}\right)^i \cdot \left(\left(\frac{1}{2}\right)^{t-2i} - \left(\frac{1}{2(1+\eta)}\right)^{t-2i} \right) \\
&= \sum_{i=\max(0, t-k/2)}^{\lfloor t/2 \rfloor} \left| (-1)^i \binom{t}{i} \cdot \left(\frac{1}{2}\right)^i \cdot \left(\frac{1}{2(1+\eta)}\right)^{t-i} \right. \\
&\quad \left. + (-1)^{t-i} \binom{t}{t-i} \cdot \left(\frac{1}{2}\right)^{t-i} \cdot \left(\frac{1}{2(1+\eta)}\right)^i \right| \\
&\leq \left| \sum_{i=0}^t (-1)^i \binom{t}{i} \cdot \left(\frac{1}{2}\right)^i \cdot \left(\frac{1}{2(1+\eta)}\right)^{t-i} \right| \\
&\leq \left(\frac{1}{2} - \frac{1}{2(1+\eta)} \right)^t \\
&= \left(\frac{\eta}{2(1+\eta)} \right)^t \leq \frac{\eta}{2(1+\eta)}.
\end{aligned}$$

Next consider the case of even $t \in [k]$. First use the binomial identity $\binom{t}{i} = \binom{t-1}{i} + \binom{t-1}{i-1}$ to obtain the following bound

$$\begin{aligned}
&\leq \left| \sum_{i=\max(0, t-k/2)}^{\min(t, k/2)} \left((-1)^{t-1-(i-1)} \binom{t-1}{i-1} - (-1)^{t-1-i} \binom{t-1}{i} \right) \right. \\
&\quad \left. \cdot \prod_{j=0}^{i-1} \frac{(k/2-j)}{(k-j)} \cdot \prod_{j=0}^{t-i-1} \frac{k/2-j}{n-k-j} \right| \\
&= \left| \sum_{i=\max(0, t-1-k/2)}^{\min(t-1, k/2)} (-1)^{t-1-i} \binom{t-1}{i} \cdot \prod_{j=0}^i \frac{(k/2-j)}{(k-j)} \cdot \prod_{j=0}^{t-i-2} \frac{k/2-j}{n-k-j} \right. \\
&\quad - \mathbf{1}[t-1 > k/2] \cdot (-1)^{t-1-k/2} \binom{t-1}{k/2} \cdot \prod_{j=0}^{k/2} \frac{(k/2-j)}{(k-j)} \cdot \prod_{j=0}^{t-k/2-2} \frac{k/2-j}{n-k-j} \\
&\quad + \sum_{i=\max(0, t-1-k/2)}^{\min(t-1, k/2)} (-1)^{t-1-i} \binom{t-1}{i} \cdot \prod_{j=0}^{i-1} \frac{(k/2-j)}{(k-j)} \cdot \prod_{j=0}^{t-i-1} \frac{k/2-j}{n-k-j} \\
&\quad \left. - \mathbf{1}[t-1 \geq k/2] \cdot (-1)^{k/2} \binom{t-1}{k/2} \cdot \prod_{j=0}^{t-2-k/2} \frac{(k/2-j)}{(k-j)} \cdot \prod_{j=0}^{k/2} \frac{k/2-j}{n-k-j} \right|.
\end{aligned}$$

First note that the second term and the fourth term always evaluate to 0. Thus we can simplify the expression as

$$\begin{aligned}
&\leq \left| \sum_{i=\max(0, t-1-k/2)}^{\min(t-1, k/2)} (-1)^{t-1-i} \binom{t-1}{i} \cdot \frac{(k/2-i)}{(k-i)} \cdot \prod_{j=0}^{i-1} \frac{(k/2-j)}{(k-j)} \right. \\
&\quad \left. \cdot \prod_{j=0}^{(t-1)-i-1} \frac{k/2-j}{n-k-j} \right| + \left| \sum_{i=\max(0, t-1-k/2)}^{\min(t-1, k/2)} (-1)^{t-1-i} \binom{t-1}{i} \right.
\end{aligned}$$

$$\cdot \frac{k/2 - (t-i-1)}{n-k-(t-i-1)} \cdot \prod_{j=0}^{i-1} \frac{(k/2-j)}{(k-j)} \cdot \prod_{j=0}^{(t-1)-i-1} \frac{k/2-j}{n-k-j} \Bigg|.$$

These two terms resembles the expression of $\frac{|\lambda_{t-1}|}{\lambda_0}$ and can be bounded by grouping terms $\binom{t-1}{i}$ and $\binom{t-1}{t-1-i}$.

$$\begin{aligned} &\leq \sum_{i=\max(0, t-1-k/2)}^{\lfloor (t-1)/2 \rfloor} \binom{t-1}{i} \cdot \prod_{j=0}^i \frac{k/2-j}{k-j} \cdot \prod_{j=0}^{i-1} \frac{k/2-j}{n-k-j} \\ &\quad \cdot \left| \prod_{j=i}^{(t-1)-i-1} \frac{k/2-j}{n-k-j} - \prod_{j=i}^{(t-1)-i-1} \frac{k/2-j-1}{k-j-1} \right| \\ &+ \sum_{i=\max(0, t-1-k/2)}^{\lfloor (t-1)/2 \rfloor} \binom{t-1}{i} \cdot \prod_{j=0}^{i-1} \frac{k/2-j}{k-j} \cdot \prod_{j=0}^i \frac{k/2-j}{n-k-j} \\ &\quad \cdot \left| \prod_{j=i}^{(t-1)-i-1} \frac{k/2-j-1}{n-k-j-1} - \prod_{j=i}^{(t-1)-i-1} \frac{k/2-j}{k-j} \right| \end{aligned}$$

Using Claim 172 again yields

$$\begin{aligned} &\leq \sum_{i=\max(0, t-1-k/2)}^{\lfloor (t-1)/2 \rfloor} \binom{t-1}{i} \cdot \left(\frac{1}{2}\right)^{i+1} \cdot \left(\frac{1}{2(1+\eta)}\right)^i \\ &\quad \cdot \left(\left(\frac{1}{2}\right)^{(t-1)-2i} - \left(\frac{1}{2(1+\eta)}\right)^{(t-1)-2i} \right) \\ &+ \sum_{i=\max(0, t-1-k/2)}^{\lfloor (t-1)/2 \rfloor} \binom{t-1}{i} \cdot \left(\frac{1}{2}\right)^i \cdot \left(\frac{1}{2(1+\eta)}\right)^{i+1} \\ &\quad \cdot \left(\left(\frac{1}{2}\right)^{(t-1)-2i} - \left(\frac{1}{2(1+\eta)}\right)^{(t-1)-2i} \right) \\ &\leq \left(\frac{1}{2} + \frac{1}{2(1+\eta)}\right) \cdot \left(\frac{1}{2} - \frac{1}{2(1+\eta)}\right)^{t-1} \leq \frac{\eta}{2(1+\eta)}. \end{aligned}$$

We thus conclude that for all $t \in [k]$, $\frac{|\lambda_t|}{\lambda_0} \leq \frac{\eta}{2(1+\eta)}$. ◀

Proof of Claim 172. We prove the statement by induction on s . The case of $s = m$ holds trivially. Suppose the statement holds for all $s > s^*$, then we are going to show that the inequality also holds for $s = s^*$. Note that by the induction hypothesis

$$\begin{aligned} &\prod_{j=s^*+1}^m \frac{C-j}{2C-j} - \prod_{j=s^*+1}^m \frac{C-j}{2(1+\eta)C-j} \\ &\leq \left(\frac{C-(s^*+1)}{2C-(s^*+1)} \right)^{m-s^*} - \left(\frac{C-(s^*+1)}{2(1+\eta)C-(s^*+1)} \right)^{m-s^*}. \quad (\text{E.2}) \end{aligned}$$

To simplify the expressions we define

$$X(s) = \prod_{j=s}^m \frac{C-j}{2C-j} \quad x(s) = \frac{C-s}{2C-s}$$

and also

$$Y(s) = \prod_{j=s}^m \frac{C-j}{2(1+\eta)C-j} \quad y(s) = \frac{C-s}{2(1+\eta)C-s}.$$

Using this notation, (E.2) can be rewritten as

$$X(s^* + 1) - Y(s^* + 1) \leq x(s^* + 1)^{m-s^*} - y(s^* + 1)^{m-s^*}.$$

Rearranging the inequality we get that

$$\begin{aligned} x(s^* + 1)^{m-s^*} - X(s^* + 1) &\geq y(s^* + 1)^{m-s^*} - Y(s^* + 1) \\ x(s^*) \cdot (x(s^* + 1)^{m-s^*} - X(s^* + 1)) &\geq y(s^*) \cdot (y(s^* + 1)^{m-s^*} - Y(s^* + 1)) \\ x(s^*) \cdot x(s^* + 1)^{m-s^*} - X(s^*) &\geq y(s^*) \cdot y(s^* + 1)^{m-s^*} - Y(s^*) \\ X(s^*) - Y(s^*) &\leq x(s^*) \cdot x(s^* + 1)^{m-s^*} - y(s^*) \cdot y(s^* + 1)^{m-s^*} \end{aligned} \quad (\text{E.3})$$

Now we bound this difference

$$\begin{aligned} &\left(x(s^*)^{m-s^*+1} - y(s^*)^{m-s^*+1} \right) - \left(x(s^*) \cdot x(s^* + 1)^{m-s^*} - y(s^*) \cdot y(s^* + 1)^{m-s^*} \right) \\ &= x(s^*) \cdot \left(x(s^*)^{m-s^*} - x(s^* + 1)^{m-s^*} \right) - y(s^*) \cdot \left(y(s^*)^{m-s^*} - y(s^* + 1)^{m-s^*} \right) \end{aligned} \quad (\text{E.4})$$

We note that for fixed constants $\alpha, \beta > 1$, and $\gamma > 0$, the function $p(x) = \left(\frac{\alpha}{\beta(1+x)} \right)^\gamma - \left(\frac{\alpha-1}{\beta(1+x)-1} \right)^\gamma$ is decreasing when $x \geq 0$. If we set $\alpha = C - s^*$, $\beta = 2C - s^*$, and $\gamma = m - s^*$, then

$$x(s^*)^{m-s^*} - x(s^* + 1)^{m-s^*} = p(0), \quad \text{while} \quad y(s^*)^{m-s^*} - y(s^* + 1)^{m-s^*} = p\left(\frac{2\eta C}{2C - s^*}\right).$$

Therefore

$$x(s^*)^{m-s^*} - x(s^* + 1)^{m-s^*} \geq y(s^*)^{m-s^*} - y(s^* + 1)^{m-s^*},$$

and together with the fact that $x(s^*) \geq y(s^*)$ we can deduce that (E.4) ≥ 0 . Equivalently,

$$x(s^*)^{m-s^*+1} - y(s^*)^{m-s^*+1} \geq x(s^*) \cdot x(s^* + 1)^{m-s^*} - y(s^*) \cdot y(s^* + 1)^{m-s^*} \geq X(s^*) - Y(s^*),$$

where the second inequality holds by (E.3). Note that this inequality is exactly the claim statement for $s = s^*$. Thus we finish the induction step and conclude the proof. $\triangleleft$

F Sparsifying the Johnson complexes

While both the Johnson complexes and the RGCs over $N = 2^{n-1}$ vertices have vertex degree N , the Johnson complexes have much denser vertex links, or equivalently more 2-faces. The degree of the Johnson complex $X_{\varepsilon, n}$'s vertex links is

$$d_{link} = \binom{\varepsilon n}{\varepsilon n/2} \cdot \binom{(1-\varepsilon)n}{\varepsilon n/2} \gg N^{\varepsilon/2},$$

while that of the RGCs is $\text{polylog}(N)$. In fact we can sparsify the links and reduce the number of 2-faces in Johnson complexes by keeping each 2-face with probability p , and set p so that the link degree becomes $\text{polylog}(N)$. Then we shall use the following theorem to show that the link expansion is preserved after subsampling.

► **Theorem 173** (Theorem 1.1 [21]). *Suppose G is a graph on n vertices with second largest eigenvalue in absolute value $\lambda_2(G)$ and minimum degree $d_{\min}$. A random subgraph H of G with edge-selection probability p has a second eigenvalue $\lambda_2(H)$ satisfying*

$$\lambda_2(H) \leq \lambda_2(G) + O\left(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log \log n)^{3/2}}\right)$$

with probability $> 1 - f(n)$ where $f(n)$ tends to 0 faster than any polynomial in $\frac{1}{n}$.

From this eigenvalue preservation result we can get the following sparsification.

► **Corollary 174.** *Consider a random subcomplex X' of a 2-dimensional Johnson complex $X_{\varepsilon,n}$ constructed by selecting each 2-face i.i.d. with probability $p \geq \frac{(h(\varepsilon) \log N)^{3/2}}{d_{\text{link}}}$ where $N = 2^{n-1}$, and d_{link} is the degree of $X_{\varepsilon,n}$'s vertex links. Then with high probability every vertex link of X' is a $\frac{1}{2}(1 - \varepsilon)(1 + o_n(1))$ -spectral expander.*

Here we recall that $h(x)$ is the binary entropy function.

Proof. We first argue that almost surely $X'(1) = X_{\varepsilon,n}(1)$ so that for every $v \in X'(0)$ the vertex link X'_v has the same vertex set as $(X_{\varepsilon,n})_v$. From the subsampling process we know that for any edge $\{v, v+s\} \in X_{\varepsilon,n}(1)$, the degree of $v+s$ in the link of v follows a binomial distribution with

$$\mathbb{E}[\text{degree of } v+s \text{ in } X'_v] = p \cdot d_{\text{link}} = (h(\varepsilon) \log N)^{3/2}.$$

By the Chernoff bound for binomial distributions we know that

$$\mathbb{P}[\text{degree of } v+s \text{ in } X'_v \leq 0.1(h(\varepsilon) \log N)^{3/2}] \leq e^{-0.45(h(\varepsilon) \log N)^{3/2}} = N^{-\Theta(\sqrt{\log N})}.$$

So by a simply union bound over all pairs of $\{v, v+s\}$ we get that with probability $1 - N^{1+h(\varepsilon)} \cdot N^{-\Theta(\sqrt{\log N})} = 1 - o(1)$ every link in X'_v has the same vertex set as $(X_{\varepsilon,n})_v$.

Next, recall by Lemma 59 $X_{\varepsilon,n}$'s vertex links have second eigenvalue $< \frac{1}{2}(1 - \varepsilon)$. Since the choice of p ensures that $\sqrt{\frac{\log N^{h(\varepsilon)}}{p \cdot d_{\text{link}}}} + \frac{(\log h(\varepsilon))^{3/2}}{p \cdot d_{\text{link}}(\log \log h(\varepsilon)N)^{3/2}} = o(1)$, each link of X' is a $\frac{1}{2}(1 - \varepsilon)(1 + o(1))$ expander. Finally, applying an union bound over all links of X' completes the proofs. ◀

G Proof of the van Kampen lemma

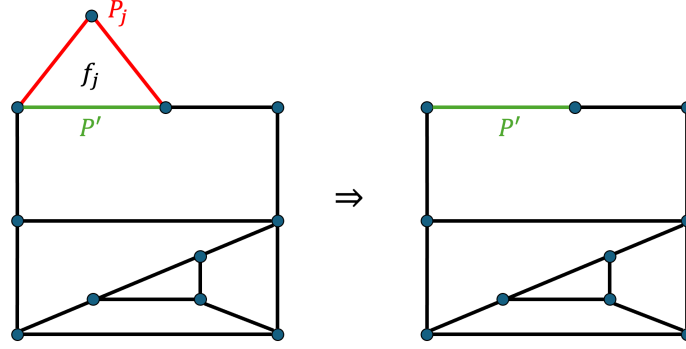
To prove this lemma, we would like to contract C_0 using the contractions of $\{C_i\}_{i>0}$ so the proof should give us an order in which C_i -s are contracted. This order is given by the order in which the set $\{f_i\}_{i>0}$ are contracted to reduce $\mathbb{R}^2 \setminus f_0$ to a point. Therefore, for every such diagram, it suffices to find a single face f_j such that we can contract to obtain a diagram of a cycle with one less interior face. The property we need from f_j is that the shared boundary of f_j and f_0 is a contiguous path. This requirement is captured in the following definition.

► **Definition 175** (valid face and path). *Let H be a van Kampen diagram. We say that an interior face f_j is valid if we can write its boundary cycle $\tilde{C}_j = P_j \circ Q$ such that the exterior cycle can be written as $\tilde{C}_0 = P_j \circ Q'$ where P_j contains at least one edge and the vertices of Q, Q' are disjoint (except for the endpoints). In this case we call P_j a valid path.*

Lemma 79 will easily follow from the following three claims.

- ▷ Claim 176. Every 2-connected plane graph H has at least one valid face.
- ▷ Claim 177. Let $P_j = (v_0, v_1, \dots, v_m)$ be a valid path in H . Then for every $1 \leq i \leq m-1$, $\deg(v_i) = 2$.
- ▷ Claim 178. Let $H = (V, E)$ be a van Kampen diagram, let $P_j = (v_0, v_1, \dots, v_m)$ and let $H_j = (V', E')$ be the induced subgraph on $V' = V \setminus \{v_1, v_2, \dots, v_{m-1}\}$. Then H_j is 2-vertex connected.

Given the three claims, we can prove the van Kampen lemma as follows.



■ **Figure 12** Contracting a face in H to obtain H_j in Lemma 79.

Proof of Lemma 79. Fix X, C_0, H, F, ψ be as defined in Section 5.3. We prove this lemma by induction on the number of inner faces of H . If H has exactly one inner face f_1 , then its boundary $\tilde{C}_1$ is also the boundary $\tilde{C}_0$ of the outer face f_0 . Then we can deduce that $C_0 = \psi(\tilde{C}_0) = \psi(\tilde{C}_1) = C_1$. So indeed there is a contraction of C_0 using $m_0 = m_1$ -triangles.

Assuming the lemma is true for all van Kampen diagrams with at most ℓ inner faces, we shall prove it for van Kampen diagrams H, F with $\ell + 1$ inner faces. By Claim 176, we can always find a valid face f_j in H such that $\tilde{C}_0 = P_j \circ Q$ and $\tilde{C}_j = P_j \circ P'$ where P_j is the corresponding valid path. The decompositions are illustrated in Figure 12.

Using Claim 85, we can contract C_0 (the image of the outer boundary of H under ψ) to $\psi(Q \circ P')$ using m_j triangles. Correspondingly, we can construct an induced subgraph H_j of H by removing all edges and intermediate vertices in P_j .

Note that the outer boundary of H_j is $\tilde{C}'_0 = Q \circ P'$. Furthermore, since the intermediate vertices in P_j all have degree 2 in H (by Claim 177), removing these vertices only removes the inner face f_j . Therefore, H_j has ℓ inner faces. Finally by Claim 178, H_j is still a 2-connected plane graph.

Thereby $H_j, F \setminus \{f_j\}$ is a van Kampen diagram of the cycle $\psi(\tilde{C}'_0)$ in X with ℓ inner faces, and so the induction hypothesis applies. Thus we complete the inductive step and the lemma is proven. ◀

Next we shall prove the three supporting claims. Before doing so, we define the cyclic ordering of vertices in a cycle which will later be used in the proofs.

► **Definition 179** (Cyclic ordering). Let $H = (V, E)$ be a graph. For a given cycle C in H , a cyclic ordering of the vertices in C is such that two consecutive vertices in the ordering share an edge in C .

Proof of Claim 176. We will show that either there exists a valid face, or that there exists two inner faces $f_i \neq f_j$ and four vertices v_1, v_2, v_3, v_4 in the cycle C_0 such that $v_1, v_3 \in f_i$, $v_2, v_4 \in f_j$ (that is, on the boundaries of the respective faces), and such that in the cyclic ordering of C_0 , $v_1 \leq v_2 \leq v_3 \leq v_4$. Then we will show that the latter case implies there is an embedding of a 5-clique on the plane, reaching a contradiction. We mention that these points do not need to be vertices in H , only points on the plane.

Indeed, take a face f_i that shares an edge with f_0 . If it is not valid, it partitions $C_0 \setminus \partial f_i$ into disconnected segments (and this partition is not empty since otherwise f_i was valid). If there exists some f_j that touches two distinct segments, we can take $v_1, v_3 \in f_i$ and $v_2, v_4 \in f_j$ as above. Otherwise, we take some $S \subseteq C_0 \setminus \partial f_i$, that contains an edge. We now search for a valid face whose path is contained in S , and continue as before. That is, take some $f_{i'}$ that shares an edge in S with the outer face. If it is valid we take $f_{i'}$ and if not we consider the partition of $S \setminus f_{i'}$ as before. We note that this process stops because the number of edges in the segment in every step strictly decreases - and when we stop we either find a valid face or four points as above.

Now let us embed the 5-clique. Indeed, we fix a point on the outer face v_0 in the interior of f_0 , and take $\{v_0, v_1, v_2, v_3, v_4\}$ to be the vertices of the clique. The edges are drawn as follows.

1. From v_0 to every other vertex we can find four non intersecting curves that only go through f_0 .²⁷
2. The edges $\{v_1, v_2\}, \{v_2, v_3\}, \{v_3, v_4\}, \{v_4, v_1\}$ are the curves induced by the cycle C_0 . These curves do not intersect the previous ones (other than in v_1, v_2, v_3, v_4) since the curves in the first item lie in the interior of f_0 .
3. By assumption, the pairs $v_1, v_3 \in f_i$ and $v_2, v_4 \in f_j$ so we can connect these pairs by curves that lie in the interiors of f_i and f_j respectively, and in particular they do not intersect any of the other curves. $\triangleleft$

Proof of Claim 177. Let f_j be the valid face. Let us denote by e_0, e_2 the edges adjacent to v in P_0 which is not an endpoint. Assume towards contradiction that there is another edge e_1 adjacent to v . Thus e_1 either lies in the interior or the exterior of f_j .

If it lies in the interior, then by 2-vertex connectivity there exists a path starting at e_1 connecting to another point in the boundary of f_j , and other than v and the end point, the path lies in the interior of f_j . This is a contradiction, since any such path disconnects f_j to two faces. Otherwise it lies in the exterior, but the argument is the same only with f_0 instead of f_j . $\triangleleft$

Proof of Claim 178. Let v_0, v_1 the end points in P_1 , and notice that v_0, v_1 participate in a cycle in H - the cycle bounding the outer face. In particular H_j is connected. Remove any vertex z from H_j . Let $P_{x,y}$ be a path that connects two vertices x, y in $H \setminus \{z\}$ (that may contain some vertices in P_j). Let us show that we can modify $P_{x,y}$ to some P' in $H_j \setminus \{z\}$. By Claim 177 every vertex in P_j which we removed had degree 2. Therefore any path containing a vertex in P_j , contains *all* of P_j (or its reverse). Thus to mend $P_{x,y}$ we just need to show that v_0, v_1 are in a simple cycle in H_j . Indeed, this is true from the definition of a valid

²⁷ For this one can use the Riemann mapping theorem to isomorphically map the outer face to the unit disc and C_0 to its boundary. Without loss of generality we can choose v_0 so that it is mapped to the center. Then we draw straight lines from the image of v_0 to the rest of the points in the boundary (and use the inverse to map this diagram back to our graph).

path: $C_j = P_j \circ Q$ and $C_0 = P_j \circ Q'$ such that all vertices in Q, Q' are disjoint besides the end points. The paths Q, Q' participate in simple cycles therefore $Q' \circ Q^{-1}$ is a simple cycle and we conclude. $\triangleleft$

H Expansion of containment walks in the Grassmann poset

To stay self contained we reprove Claim 21 in a similar manner as in [30].

Proof of Claim 21. Although all the claims below apply for Y_w for any $w \in Y$, for a simple presentation let us assume that $w = \{0\}$ (the same idea works for arbitrary w). Let us denote by $U_{i,j}$ the bipartite graph operator from $Y(i)$ to $Y(j)$. We note that $U_{i,j} = U_{j-1,j} \dots U_{i+1,i+2} U_{i,i+1}$, and therefore it is sufficient to prove that $\lambda_2(U_{i,i+1}) \leq (0.61 + i\lambda)^{\frac{1}{2}}$. We do so by induction on i .

In fact, we prove a slightly stronger statement, that $\lambda_2(U_{i,i+1}) = \left(\sum_{\ell=1}^i \frac{1}{2^{\ell+1}-1} + i\lambda \right)^{\frac{1}{2}}$. This suffices because $\sum_{\ell=1}^i \frac{1}{2^{\ell+1}-1} \leq \sum_{\ell=1}^{\infty} \frac{1}{2^{\ell+1}-1} \leq 0.61$. Let $D_{i+1,i}$ be the adjoint operator to $U_{i,i+1}$, and denote by $\lambda_i = \lambda(D_{i+1,i} U_{i,i+1})$ be the second largest eigenvalue of the two step walk from $Y(i)$ to $Y(i)$. To show that $\lambda_2(U_{i,i+1}) = \left(\sum_{\ell=1}^i \frac{1}{2^{\ell+1}-1} + i\lambda \right)^{\frac{1}{2}}$ it suffices to prove that $\lambda_i = \left(\sum_{\ell=1}^i \frac{1}{2^{\ell+1}-1} + i\lambda \right)$.

The base case for the induction is for $U_{1,2}$. In this case

$$\lambda_1 = \lambda(D_{2,1} U_{1,2}) = \frac{1}{3}I + \frac{2}{3}M^+$$

where M^+ is the non-lazy version of the walk (namely, where we traverse from w_1 to w_2 if $w_1 + w_2 \in Y(2)$). The adjacency operator M^+ is the link of $\{0\}$, and thus by λ -expansion we have that $\lambda_1 \leq \frac{1}{3} + \frac{2}{3}\lambda \leq \frac{1}{2^2-1} + \lambda$.

Assume that $\lambda_2(U_{i,i+1}) \leq \left(\sum_{\ell=1}^i \frac{1}{2^{\ell+1}-1} + i\lambda \right)^{\frac{1}{2}}$ and consider the two step walk on $Y(i+1)$. As above, we note that

$$\lambda_{i+1} = \lambda(D_{i+2,i+1} U_{i+1,i+2}) = \frac{1}{2^{i+2}-1}I + \left(1 - \frac{1}{2^{i+2}-1} \right) M_{i+1}^+$$

where M^+ is the non-lazy version of the walk. We note that the laziness probability $\frac{1}{2^{i+2}-1}$ is because every subspace $w \in Y(i+2)$ has exactly $2^{i+2} - 1$ subspaces of codimension 1 (so this is the regularity of the vertex). Thus we will prove below that $\lambda(M_{i+1}^+) \leq \lambda + \lambda_i \leq \left(\sum_{\ell=1}^i \frac{1}{2^{\ell+1}-1} + (i+1)\lambda \right)$ and this will complete the induction.

Let us decompose the non-lazy walk M^+ via Lemma 15. Our components are $T = \{Y_W\}_{W \in Y(i)}$. To be more explicit, note that $W_1 \sim W_2$ in the two step walk if and only if $W_1 + W_2 \in Y(i+2)$. We can choose such a pair by first choosing $W \in Y(i)$ and then choosing an $\{W_1, W_2\}$ in the link Y_W . The distribution of the link is the conditional distribution, hence taking μ to be the $Y(i)$ -marginal of the distribution over flags in Y , we have that T is a local-to-global decomposition. The local-to-global graph $\mathcal{H}_\tau$ connects $W \in Y(i)$ to $W \in Y(i+1)$ - so this is just the containment graph $(Y(i), Y(i+1))$ which by induction is a $\sqrt{\lambda_i}$ -expander. By Lemma 15, $\lambda(M_{i+1}^+) \leq \lambda + \lambda_i \leq \left(\sum_{\ell=1}^i \frac{1}{2^{\ell+1}-1} + (i+1)\lambda \right)$ and we can conclude. $\triangleleft$

I A degree lower bound

A classical result in graph theory says that a connected Cayley graph over $\mathbb{F}_2^n$ has degree at least n . Moreover, if the Cayley graph is a λ -expander then it has degree at least $\Omega(\frac{n}{\lambda^2 \log(1/\lambda)})$ [2]. Are there sharper lower bounds on the degree if the Cayley graph is also a Cayley local spectral expander?

In this section we consider the “inverse” point of view on this question. Given a graph G with m vertices, what is the maximal n such that there exists a connected Cayley complex over $\mathbb{F}_2^n$ whose link is isomorphic to G . Using this point of view, we prove new lower bounds for Cayley local spectral expanders, that depend on properties of G . Specifically, we prove that if G has no isolated vertices then $m > 1.5(n - 1)$; if G is connected we prove that $m > 2n - 1$. Finally, if the link is a d -regular edge expander, we prove a bound of $\Omega(d^{1/2}n)$. For all these bounds, we use the code theoretic interpretation of the links given in [50].

We begin by describing the link of a Cayley complex, and using this description we give a lower bound on the degree of the Cayley complex in terms of the rank of some matrix. This linear algebraic characterization of a degree lower bound is already implicit in [50, Section 6].

After that, we will lower bound the rank of the matrix, and through this we will get lower bounds on the degree of such Cayley complexes.

A link in a Cayley complex X naturally comes with labels on the edges which are induced by the generating set (i.e. an edge between $x + s_1$ and $x + s_2$ in the link of x will be labeled $s_1 + s_2$ which by assumption is a generator in the generating set). Note that if $x + s_1, x + s_2$ is labeled $s_1 + s_2$ then there is a 3-cycle in the link between $x + s_1, x + s_2$ and $x + s_1 + s_2$ where every edge whose endpoints correspond to two generators is labeled by the third generator.

Let $G = (V, E)$ be a graph with vertex set $V = \{v_1, v_2, \dots, v_m\}$ and a labeling $\ell : E \rightarrow V$. The graph G is *nice* if for every $\{v_i, v_j\}$ that is labeled v_k there is a three cycle v_i, v_j, v_k in the graph where $\{v_i, v_k\}$ is labeled v_j and $\{v_j, v_k\}$ is labeled v_i . We would like to understand for which n does there exist a Cayley complex X over $\mathbb{F}_2^n$ so that the vertex links of X are isomorphic to G . That is, for which n can we find generators $s_1, s_2, \dots, s_m \in \mathbb{F}_2^n$ such that the map $\phi(s_i) = v_i$ is a graph isomorphism between the vertex link and G . Observe that for this to hold, whenever there is an edge $\{v_i, v_j\} \in E$ labeled by v_k , we require that $s_i + s_j + s_k = 0$. Thus we can define the matrix $H_G \in \mathbb{F}_2^{|E| \times |V|}$ to have in every row the indicator vector of i, j, k for every $\{v_i, v_j\} \in E$ labeled by v_k . Let us show that this is a sufficient condition.

▷ **Claim 180.** Let G be a nice graph with m vertices. Let $S = \{s_1, s_2, \dots, s_m\} \in \mathbb{F}_2^n$ and let M_s be the matrix whose rows are the s_i 's. Then the following is equivalent.

1. The columns of M_s are in the kernel of H_G .
2. There exists a Cayley complex X over $\mathbb{F}_2^n$ and whose link is isomorphic to G whose underlying graph is $\text{Cay}(\mathbb{F}_2^n, S)$.

In particular, if X is connected then $m \geq n + \text{rank}(H_G)$.

This equivalence is essentially what was proven in [50, Lemma 77]. The degree bound is a direct consequence of it. Nevertheless, we prove it in our language for clarity.

Proof. If the first item occurs, this means that for every $\{v_i, v_j\} \in E$ labeled by v_k , $s_i + s_j + s_k = 0$ and therefore the 3-cycles $(x, x + s_i, x + s_i + s_j = x + s_k, x)$ are in $\text{Cay}(\mathbb{F}_2^n, S)$ so add $\{x + s_i, x + s_j, x + s_k\}$ into the triangle set of X . Then it is straightforward to verify that the link of X is isomorphic to G .

On the other hand, from the discussion above, the isomorphism to G implies that for every edge $\{v_i, v_j\}$ in G with label, their corresponding generators satisfy $s_i + s_j + s_k = 0$. Hence the columns of M_s are in $\text{Ker}(H_G)$.

In particular, this implies that $\text{rank}(M_s) \leq \dim(\text{Ker}(H_G)) = m - \text{rank}(H_G)$. On the other hand, if the complex is connected then $\text{rank}(M_s) \geq n$. Combining the two we get that the degree of X is lower bounded by $m \geq n + \text{rank}(H_G)$. $\triangleleft$

► **Corollary 181.** *Let X be a connected Cayley complex over $\mathbb{F}_2^n$ with degree m such that the link of X contains no isolated vertices. Then $m \geq 1.5(n - 1)$. Moreover, if the link is connected then $m \geq 2n - 1$.*

Proof. The corollary follows from the inequality $m \geq n + \text{rank}(H_G)$, together with the observation that $\text{rank}(H_G) \geq \frac{m}{3}$ which we now explain. We can greedily select independent rows from H_G such that whenever a generator s_i has not appeared in any selected rows, we take an row where s_i appears. As every newly selected row contains at most 3 new generators, we can find at least $\lfloor \frac{m}{3} \rfloor$ such equations. Thus, $m \geq n + \lfloor \frac{m}{3} \rfloor$ and the bound follows.

If the link is connected, we can improve on this argument. At every step when we select a row, we denote by $D \subseteq V$ the set of vertices that have appeared in the selected rows. Since the link is connected, therefore while $D \neq V$ we can always choose a new row containing vertices from both D and its complement, in which case the new row (besides the first one) introduces at most 2 new vertices. Thus $\text{rank}(H_G) \geq 1 + \frac{m-3}{2} = \frac{m}{2} - \frac{1}{2}$. The bound follows. $\blacktriangleleft$

1.1 Improved bound from expansion

If the graph G is an (edge) expander, we can improve the naive bounds above. Let m be the degree of the Cayley complex and d be the degree of a vertex in the link. We get a degree bound of $m = \Omega(d^{1/2}n)$. In particular, if the degree goes to infinity (which is also a necessary condition to get better and better spectral expansion), this lower bound tends to infinity as well.

► **Lemma 182.** *Let X be a simple Cayley HDX over $\mathbb{F}_2^n$ with m generators.²⁸ Assume that the link is d -regular and denote it by $G = (V, E)$. Let $c \in (0, 1)$ be such that for every subset of vertices $D \subseteq V$ in the link ,*

$$\mathbb{P}_{(x,y) \sim E} [x \in D, y \notin D] \geq c \mathbb{P}_{x \sim V} [x \in D] \mathbb{P}_{y \sim V} [y \in V \setminus D]$$

where the probability is over sampling an edge (oriented) in the link. Then

$$m = \Omega(cd^{1/2}n).$$

A few remarks are in order.

► **Remark 183.**

1. We note that the edge expansion condition in the lemma is equivalent to one-sided λ -spectral expansion for some $\lambda < 1$ by Cheeger's inequality. Thus whenever λ is bounded away from 1 we can treat c as some constant absorbed in the Ω notation, and write $m = \Omega(d^{1/2}n)$.

²⁸ Simple, means that there are no loops or multi-edges. This also implies that every two labels of edges that share a vertex are distinct.

2. Moreover, the well-known Alon-Boppana bound implies that d -regular λ -one sided spectral expanders have $d = \Omega(\lambda^{-2})$. Therefore, this bound also implies

$$m = \Omega(\lambda^{-1}n).$$

However, this is not as sharp as the bound given in [2].

3. We prove this lemma for regular graphs for simplicity. The argument could be extended for more general edge expanders.

Proof of Lemma 182. We will assume without loss of generality that $d \geq 36$ since otherwise the claim is trivial from the bound $m \geq n$. By Claim 180, $m \geq n + \text{rank}(H_G)$ therefore it suffices to prove that

$$\text{rank}(H_G) \geq \left(1 - O\left(\frac{1}{cd^{1/2}}\right)\right) m \quad (\text{I.1})$$

to prove the lemma.

This lower bound follows from the same greedy row selection process we conducted above. We start by greedily selecting rows from H_G such that the selected rows are independent. We keep track of the vertices/variables that appear in the selected rows, and select new rows such that every new row introduces the least number of new vertices possible. We will prove that there is a way to select rows so that only $O(\frac{1}{d^{1/2}c})$ fraction of the rows selected will introduce more than one new vertex (and the number of newly introduced vertices in one step is always ≤ 3). Since the process terminates after $\text{rank}(H_G)$ steps, the argument above implies that $\text{rank}(H_G) \cdot (1 + O(\frac{1}{d^{1/2}c})) \geq m$ which in turn implies (I.1).

For $i = 1, 2, \dots, \text{rank}(H_G)$ we denote by D_i the set of variables that are contained in one of the first i rows selected. By assumption, $\mathbb{P}_{(x,y) \sim E}[x \in D_i, y \notin D_i] \geq c \mathbb{P}[x \in D_i] \mathbb{P}[y \in V \setminus D_i]$. Equivalently,

$$\mathbb{P}_{(x,y) \sim E}[x \in D_i \mid y \notin D_i] \geq c \mathbb{P}[x \in D_i].$$

In particular, for every D_i there exists a vertex $y_i \in V \setminus D_i$ such that at least $c \mathbb{P}[x \in D_i] \cdot d$ of its outgoing edges go into D_i . In particular, once $\mathbb{P}[x \in D_i] \geq \frac{1}{cd^{1/2}}$ every y_i has at least $d^{1/2}$ edges into D_i .

Hence, we select rows as follows. If $\mathbb{P}[D_i] < \frac{1}{cd^{1/2}}$ we select new rows arbitrarily that introduce as few new vertices as possible. After $\mathbb{P}[x \in D_i] \geq \frac{1}{cd^{1/2}}$, if in the i -th step every new row introduces at least two new vertices, we select a row that contains y_i (as defined above) and a vertex in D_i (such a row exists by our choice of y_i). We observe that if every new row introduces at least two new vertices, this implies that the edges connecting y_i to D_i are labeled by distinct generators in $V \setminus D_i$. In particular, after selecting y_i which adds y_i and another new vertex to D_{i+1} , there are at least $d^{1/2} - 1 \geq \frac{1}{3}d^{1/2}$ edges from y_i to D_{i+1} that are labeled by distinct vertices in $V \setminus D_{i+1}$ (since these edge labels were not in D_i when we inserted y_i). So the next $\geq \frac{1}{3}d^{1/2}$ steps of the process add at most 1 new vertex per step. In particular, for every such step where 2 vertices are inserted to D_i to get D_{i+1} , there is a sequence of $\geq \frac{1}{3}d^{1/2}$ steps each of which adds ≤ 1 new vertex. In conclusion, there are at most $O(\frac{3m}{d^{1/2}} + \frac{m}{2cd^{1/2}})$ steps that introduce more than one new vertex to D_i (out of at least $\frac{m}{2} - \frac{1}{2}$ steps) and the lemma follows. $\blacktriangleleft$