Almost Euclidean subspaces of ℓ_1^N via expander codes*

Venkatesan Guruswami† James R. Lee‡ Alexander Razborov§

February, 2009

Abstract

We give an explicit (in particular, deterministic polynomial time) construction of subspaces $X \subseteq \mathbb{R}^N$ of dimension $(1 - o(1))N$ such that for every $x \in X$,

$$(\log N)^{-O(\log\log\log N)}\sqrt{N}\,\|x\|_2 \leqslant \|x\|_1 \leqslant \sqrt{N}\,\|x\|_2.$$

If we are allowed to use $N^{1/\log\log N} \leqslant N^{o(1)}$ random bits and $\mathsf{dim}(X) \geqslant (1-\eta)N$ for any fixed constant η, the lower bound can be further improved to $(\log N)^{-O(1)}\sqrt{N}\|x\|_2$.

Through known connections between such Euclidean sections of ℓ_1 and compressed sensing matrices, our result also gives explicit compressed sensing matrices for low compression factors for which basis pursuit is guaranteed to recover sparse signals. Our construction makes use of unbalanced bipartite graphs to impose local linear constraints on vectors in the subspace, and our analysis relies on expansion properties of the graph. This is inspired by similar constructions of error-correcting codes.

Mathematics Subject Classification (2000) codes: 68R05, 68P30, 51N20.

Abbreviated title: Explicit Euclidean sections from expander codes.

---

*A preliminary version of this paper appeared in the *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, January 2008.

†**Corresponding author.** Affiliation and address: University of Washington, Department of Computer Science and Engineering, Box 352350, Seattle, WA 98195. Part of this work was done when the author was on leave at the School of Mathematics, Institute for Advanced Study, Princeton, NJ. Research supported in part by NSF CCF-0343672, a Packard Fellowship, and NSF grant CCR-0324906 to the IAS. `venkat@cs.washington.edu`

‡University of Washington, Department of Computer Science and Engineering, Seattle, WA 98195. Research supported in part by NSF CCF-0644037. `jrl@cs.washington.edu`

§Institute for Advanced Study, School of Mathematics, Princeton, NJ and Steklov Mathematical Institute, Moscow, Russia. Current address: University of Chicago, Department of Computer Science, Chicago, IL 60637. `razborov@cs.uchicago.edu`

1 Introduction

Classical results in high-dimensional geometry [13, 23] state that a random (with respect to the Haar measure) subspace $X \subseteq \mathbb{R}^N$ of dimension εN [13] or even $(1-\varepsilon)N$ [23] is an almost Euclidean section in ℓ_1^N, in the sense that $\sqrt{N}\|x\|_1$ and $\|x\|_2$ are within constant factors, uniformly for every $x \in X$. Indeed, this is a particular example of the use of the probabilistic method, a technique which is now ubiquitous in asymptotic geometric analysis.

On the other hand, it is usually the case that objects constructed in such a manner are very hard to come by *explicitly.* Motivated in part by ever growing connections with combinatorics and theoretical computer science, the problem of explicit constructions of such subspaces has gained substantially in popularity over the last several years; see, e.g. [36, Sec. 4], [30, Prob. 8], [22, Sec. 2.2]. Indeed, such subspaces (viewed as embeddings) are important for problems like high-dimensional nearest-neighbor search [19] and compressed sensing [10], and one expects that explicit constructions will lead, in particular, to a better understanding of the underlying geometric structure. (See also the end of the introduction for a discussion of the relevance to compressed sensing.)

1.1 Previous results and our contributions

If one relaxes the requirement that $\mathsf{dim}(X) = \Omega(N)$ or allows a limited amount of randomness in the construction, a number of results are known. In order to review these, we define the *distortion* $\Delta(X)$ *of* $X \subseteq \mathbb{R}^N$ by

$$\Delta(X) = \sqrt{N} \cdot \max_{0 \neq x \in X} \frac{\|x\|_2}{\|x\|_1}.$$

In the first direction, it is well-known that an explicit construction with distortion $O(1)$ and $\mathsf{dim}(X) = \Omega(\sqrt{N})$ can be extracted from Rudin [32] (see also [26] for a more accessible exposition). Indyk [20] presented a deterministic polynomial-time construction with distortion $1 + o(1)$ and $\mathsf{dim}(X) \geqslant \frac{N}{\exp(O(\log\log N)^2)}$.

Another very interesting line of research pursued by various authors and in quite different contexts is to achieve, in the terminology of theoretical computer science, a *partial derandomization* of the original (existential) results. The goal is to come up with a "constructive" *discrete* probabilistic measure on subspaces X of $\mathbb{R}^N$ such that a random (with respect to this measure) subspace still has low distortion almost surely, whereas the entropy of this measure (that is, the number of truly random bits necessary to sample from it) is also as low as possible.

Denoting by $\mathcal{A}_{k,N}$ a random $k \times N$ sign matrix (i.e. with i.i.d. Bernoulli ± 1 entries), one can extract from the paper [23] by Kashin that $\ker(\mathcal{A}_{k,N})$, a subspace of codimension at most k has, with high probability, distortion $\sqrt{N/k} \cdot \mathrm{polylog}(N/k)$. Schechtman [33] arrived at similar conclusions for subspaces generated by rows of $\mathcal{A}_{N-k,N}$. Artstein-Avidan and Milman [2] considered again the model $\ker(\mathcal{A}_{k,N})$ and derandomized this further from $O(N^2)$ to $O(N \log N)$ bits of randomness. We remark that the pseudorandom generator

approach of Indyk [19] can be used to efficiently construct such subspaces using $O(N \log^2 N)$ random bits. This was further improved to $O(N)$ bits by Lovett and Sodin [27]. Subsequent to our work, Guruswami, Lee, and Wigderson [16] used the construction approach from this paper to reduce the random bits to $O(N^\delta)$ for any $\delta > 0$ while achieving distortion $2^{O(1/\delta)}$.

As far as deterministic constructions with $\mathsf{dim}(X) = \Omega(N)$ are concerned, we are aware of only one result; implicit in various papers (see e.g. [11]) is a subspace with $\mathsf{dim}(X) = N/2$ and distortion $O(N^{1/4})$. For $\mathsf{dim}(X) \geqslant 3N/4$, say, it appears that nothing non-trivial was shown prior to our work.

Our main result is as follows.

Theorem 1.1. *For every $\eta = \eta(N)$, there is an explicit, deterministic polynomial-time construction of subspaces $X \subseteq \mathbb{R}^N$ with $\mathsf{dim}(X) \geqslant (1-\eta)N$ and distortion $(\eta^{-1} \log\log N)^{O(\log\log N)}$.*

Like in [23, 2, 27], our space X has the form $\ker(A_{k,N})$ for a sign matrix $A_{k,N}$, but in our case this matrix is completely explicit (and, in particular, polynomial time computable). Its high-level overview is given in Section 1.2.3 below.

On the other hand, if we allow ourselves a small number of random bits, then we can slightly improve the bound on distortion.

Theorem 1.2. *For every fixed $\eta > 0$ there is a polynomial time algorithm using $N^{1/\log\log N}$ random bits that almost surely produces a subspace $X \subseteq \mathbb{R}^N$ with $\mathsf{dim}(X) \geqslant (1-\eta)N$ and distortion $(\log N)^{O(1)}$.*

1.2 Proof techniques

1.2.1 Spreading subspaces

Low distortion of a section $X \subseteq \mathbb{R}^N$ intuitively means that for every non-zero $x \in X$, a "substantial" portion of its mass is spread over "many" coordinates, and we formalize this intuition by introducing the concept of a *spread subspace* (Definition 2.10). While this concept is tightly related to distortion, it is far more convenient to work with. In particular, using a simple spectral argument and Kerdock codes [25], [29, Chap. 15], we initialize our proof by presenting explicit subspaces with reasonably good spreading properties. These codes appeared also in the approach of Indyk [20], though they were used in a dual capacity (i.e., as generator matrices instead of check matrices). In terms of distortion, however, this construction can achieve at best $O(N^{1/4})$.

1.2.2 The main construction

The key contribution of our paper consists in exploiting the natural analogy between low-distortion subspaces over the reals and error-correcting codes over a finite alphabet.

Let $G = (\{1, 2, \ldots, N\}, V_R, E)$ be a bipartite graph which is d-regular on the right, and let $L \subseteq \mathbb{R}^d$ be any subspace. Using the notation $\Gamma(j) \subseteq \{1, 2, \ldots, N\}$ for the neighbor set of a vertex $j \in V_R$, we analyze the subspace

$$X(G, L) = \{x \in \mathbb{R}^N : x_{\Gamma(j)} \in L \text{ for every } j \in V_R\},$$

where for $S \subseteq [N]$, $x_S \in \mathbb{R}^{|S|}$ represents the vector x restricted to the coordinates lying in S. In other words, we impose local linear constraints (from L) according to the structure of some bipartite graph G. As Theorem 4.2 shows, one can in particular analyze the spreading properties of $X(G, L)$ in terms of those of L and the expansion properties of G.

1.2.3 Putting it together: combinatorial overview

Our final space X will be of the form $X = \bigcap_{i=0}^{r-1} X(G_i, L_i)$ for suitably chosen G_i, L_i (see the proof of Theorem 1.1). Combinatorially this simply means that we take $k_i \times N$ sign matrices A_i such that $X(G_i, L_i) = \ker(A_i)$ and stack them on the top of one another to get our final matrix $A_{k,N}$. Moreover, every A_i is a stack of $|V_R|$ copies of the sign matrix A'_i with $\ker(A'_i) = L_i$ in which every copy is padded with $(N - d)$ zero columns. The exact placement of these columns is governed by the graph G that is chosen to satisfy certain expansion properties (Theorem 2.6), and it is different in different copies.

And then we have one more level of recursion: Every L_i has the form $X(G'_i, L'_i)$, where G'_i again have certain (but this time different – see Proposition 2.7) expansion properties and L'_i is our initial subspace (see Section 1.2.1).

1.2.4 Connections to discrete codes

Our approach is inspired by *Low Density Parity Check Codes* (LDPC) introduced by Gallager [14]. They are particularly suited to our purposes since, unlike most other explicit constructions in coding theory, they exploit a *combinatorial* structure of the parity check matrix and rely very little on the arithmetic of the underlying finite field. Sipser and Spielman [35] showed that one can achieve basically the same results (that is, simple and elegant constructions of constant rate, constant relative minimal distance codes) by considering adjacency matrices of sufficiently good expanders instead of a *random* sparse matrix. These codes are nowadays called *expander codes*. Using an idea due to Tanner [37], it was shown in [35] (see also [39]) that even better constructions can be achieved by replacing the parity check by a small (constant size) inner code. Our results demonstrate that analogous constructions work over the reals: If the inner subspace L has reasonably good spreading properties, then the spreading properties of $X(G, L)$ are even better. Upper bounds on distortion follow.

1.3 Organization

In Section 2, we provide necessary background on bipartite expander graphs and define spread subspaces. In Section 3, we initialize our construction with an explicit subspace with reasonably good spreading properties. In Section 4, we describe and analyze our main expander-based construction. Finally, in Section 5, we discuss why improvements to our bounds may have to come from a source other than better expander graphs.

1.4 Relationship to compressed sensing.

In [9], DeVore asks whether probabilistically generated compressed sensing matrices can be given by deterministic constructions.

The note [24] makes the connection between distortion and compressed sensing quite explicit. If $M : \mathbb{R}^N \to \mathbb{R}^n$ satisfies $\Delta(\ker(M)) \leqslant D$, then any vector $x \in \mathbb{R}^N$ with $|\mathrm{supp}(x)| < \frac{N}{4D^2}$ can be uniquely recovered from its encoding Mx. Moreover, given the encoding $y = Mx$, the recovery can be performed efficiently by solving the following convex optimization problem: $\min_{v \in \mathbb{R}^N} \|v\|_1$ subject to $Mv = y$.

In fact, something more general is shown. Define, for $x \in \mathbb{R}^N$, the quantity

$$\sigma_k(x)_1 = \min_{w \in \mathbb{R}^N : |\mathrm{supp}(w)| \leqslant k} \|x - w\|_1 \tag{1}$$

as the error of the best sparse approximation to x. Then given Mx, the above algorithm recovers a vector $v \in \mathbb{R}^N$ such that $Mx = Mv$ and $\|x - v\|_2 \leqslant \frac{\sigma_k(x)_1}{\sqrt{k}}$, for $k = \Theta(N/D^2)$. In other words, the recovery algorithm is *stable* in the sense that it can also tolerate noise in the signal x, and is able to perform approximate recovery even for signals which are only approximately sparse.

Thus our results show the existence of a mapping $M : \mathbb{R}^N \to \mathbb{R}^{o(N)}$, where M is given by an explicit matrix, and such that any vector $x \in \mathbb{R}^N$ with $|\mathrm{supp}(x)| \leqslant \frac{N}{(\log N)^{C \log\log\log N}}$ can be efficiently recovered from Mx (the stable generalization also holds, along the lines of (1)). This yields the best-known explicit compressed sensing matrices for this range of parameters (e.g. where $n \approx N/\mathrm{poly}(\log N)$). Moreover, unlike probabilistic constructions, our matrices are quite sparse, making compression (i.e., matrix-vector multiplication) and recovery (via Basis Pursuit) more efficient. For instance, when $n = N/2$, our matrices have only $N^{2-\varepsilon}$ non-zero entries for some $\varepsilon > 0$. We refer to [21] for explicit constructions that achieve a better tradeoff for $n \approx N^\delta$, with $0 < \delta < 1$. We remark that the construction of [21] is not stable in the sense discussed above (and hence only works for actual sparse signals).

2 Preliminaries

2.1 Notation

For two expressions A, B, we sometimes write $A \gtrsim B$ if $A = \Omega(B)$, $A \lesssim B$ if $A = O(B)$, and we write $A \approx B$ if $A = \Theta(B)$, that is $A \gtrsim B$ and $B \gtrsim A$. For a positive integer M, $[M]$ denotes the set $\{1, 2, \ldots, M\}$. The set of nonnegative integers is denoted by $\mathbb{N}$.

2.2 Unbalanced bipartite expanders

Our construction is based on unbalanced bipartite graphs with non-trivial vertex expansion.

Definition 2.1. *A bipartite graph $G = (V_L, V_R, E)$ (with no multiple edges) is said to be an* (N, n, D, d)-right regular graph *if $|V_L| = N$, $|V_R| = n$, every vertex on the left hand size V_L has degree* at most *D, and every vertex on the right hand side V_R has degree* equal *to d.*

For a graph $G = (V, E)$ and a vertex $v \in V$, we denote by $\Gamma_G(v)$ the *vertex neighborhood* $\{u \in V \mid (v, u) \in E\}$ of v. We denote by $d_v = |\Gamma_G(v)|$ the degree of a vertex v. The *neighborhood* of a subset $S \subseteq V$ is defined by $\Gamma_G(S) = \bigcup_{v \in S} \Gamma_G(v)$. When the graph G is clear from the context, we may omit the subscript G and denote the neighborhoods as just $\Gamma(v)$ and $\Gamma(S)$.

Definition 2.2 (Expansion profile)**.** *The* expansion profile *of a bipartite graph $G = (V_L, V_R, E)$ is the function $\Lambda_G : (0, |V_L|] \to \mathbb{N}$ defined by*

$$\Lambda_G(m) = \min \{|\Gamma_G(S)| : S \subseteq V_L, |S| \geqslant m\} .$$

Note that $\Lambda_G(m) = \min_{v \in V_L} d_v$ for $0 < m \leqslant 1$.

For our work, we need *unbalanced* bipartite graphs with expansion from the larger side to the smaller side. Our results are based on two known explicit constructions of such graphs. The first one is to take the edge-vertex incidence graph of a non-bipartite *spectral expander*[1] such as a Ramanujan graph. These were also the graphs used in the work on expander codes [35, 39]. The second construction of expanders is based on a suggestion due to Avi Wigderson. It uses a result of Barak, et. al. [4] based on sum-product estimates in finite fields; see [38, §2.8] for background on such estimates.

For our purposes, it is also convenient (but not strictly necessary) to have bipartite graphs that are regular on the right. We begin by describing a simple method to achieve right-regularity with minimal impact on the expansion and degree parameters, and then turn to stating the precise statements about the two expander constructions we will make use of in Section 4 to construct our explicit subspaces.

2.2.1 Right-regularization

Lemma 2.3. *Given a graph $H = (V_L, V_R, E)$ with $|V_L| = N$, $|V_R| = n$, that is left-regular with each vertex in V_L having degree D, one can construct in $O(ND)$ time an $(N, n', 2D, d)$-right regular graph G with $n' \leqslant 2n$ and $d = \lceil \frac{ND}{n} \rceil$ such that the expansion profiles satisfy $\Lambda_G(m) \geqslant \Lambda_H(m)$ for all $m > 0$.*

Proof. Let $d_{\mathrm{av}} = ND/n$ be the average right degree of the graph H and let $d = \lceil d_{\mathrm{av}} \rceil$. Split each vertex $v \in V_R$ of degree d_v into $\lfloor d_v/d \rfloor$ vertices of degree d each, and if $d_v \mod d > 0$, a "remainder" vertex of degree $r_v = d_v \mod d$. Distribute the d_v edges incident to v to these split vertices in an arbitrary way. The number of newly introduced vertices is at most $\sum_{v \in V_R} d_v/d = nd_{\mathrm{av}}/d \leqslant n$, so the number n' of right-side vertices in the new graph satisfies $n' \leqslant 2n$.

All vertices except the at most n "remainder" vertices now have degree exactly d. For each $v \in V_R$, add $d - r_v$ edges to the corresponding remainder vertex (if one exists). Since this step adds at most $(d-1)n \leqslant d_{\mathrm{av}} n = ND$ edges, it is possible to distribute these edges

[1] That is, a regular graph with a large gap between the largest and second largest eigenvalue of its adjacency matrix.

in such a way that no vertex in V_L is incident on more than D of the new edges. Therefore, the maximum left-degree of the new graph is at most $2D$.

The claim about expansion is obvious — just ignore the newly added edges, and the splitting of vertices can only improve the vertex expansion. □

2.2.2 Spectral expanders

The next theorem converts non-bipartite expanders to unbalanced bipartite expanders via the usual edge-vertex incidence construction.

Theorem 2.4. *For every $d \geqslant 5$ and $N \geqslant d$, there exists an explicit $(N' = \Theta(N),\ n, 2, \Theta(d))$-right regular graph G whose expansion profile satisfies $\Lambda_G(m) \geqslant \min\left\{\frac{m}{2\sqrt{d}}, \frac{\sqrt{2mN'}}{d}\right\}$.*

Proof. Let p, q be any two primes which are both congruent to 1 modulo 4. Then there exists an explicit $(p+1)$-regular graph $Y = (V, F)$ with $\frac{q(q^2-1)}{4} \leqslant |V| \leqslant \frac{q(q^2-1)}{2}$ and such that $\lambda_2 = \lambda_2(Y) \leqslant 2\sqrt{p}$, where $\lambda_2(Y)$ is the second largest eigenvalue (in absolute value) of the adjacency matrix of Y [28]. (See [18, §2] for a discussion of explicit constructions of expander graphs.)

Letting $n = |V|$, we define a $(\frac{(p+1)n}{2}, n, 2, p+1)$-right regular bipartite graph $G = (V_L, V_R, E)$ where $V_L = F$, $V_R = V$, and $(e, v) \in E$ if v is an endpoint of $e \in F$. To analyze the expansion properties of G, we use the following lemma of Alon and Chung [1].

Lemma 2.5. *If Y is any d-regular graph on n vertices with second eigenvalue λ_2, then the induced subgraph on any set of γn vertices in Y has at most*

$$\left(\gamma^2 + \gamma\frac{\lambda_2}{d}\right)\frac{dn}{2}$$

edges.

In particular, if $S \subseteq V_L$ satisfies $|S| \geqslant \gamma^2(p+1)n$ and $|S| \geqslant 2\gamma n\sqrt{p+1}$, then $|\Gamma_G(S)| \geqslant \gamma n$. Stated different, for any $S \subseteq V_L$, we have

$$|\Gamma_G(S)| \geqslant \frac{\min\left\{2\sqrt{|S|n}, |S|\right\}}{2\sqrt{p+1}}$$

Setting $N' = \frac{(p+1)n}{2}$, we see that $\Lambda_G(m) \geqslant \min\left\{\frac{m}{2\sqrt{d}}, \frac{\sqrt{2N'm}}{d}\right\}$.

Now given parameters $d \geqslant 5$ and $N \geqslant d$, let p be the largest prime satisfying $p + 1 \leqslant d$ and $p \equiv 1 \pmod 4$, and let q be the smallest prime satisfying $\frac{q(q^2-1)(p+1)}{8} \geqslant N$ and $q \equiv 1 \pmod 4$. The theorem follows by noting that for all integers $m \geqslant 3$, there exists a prime $p \in [m, 2m]$ which is congruent to 1 modulo 4 (see [12]). □

The expanders of Theorem 2.4 are already right-regular but they have one drawback; we cannot fully control the number of left-side vertices N. Fortunately, this can be easily circumvented with the same Lemma 2.3.

Theorem 2.6. *For every $d \geqslant 5$ and $N \geqslant d$, there exists an explicit $(N, n, 4, \Theta(d))$-right regular graph G which satisfies $\Lambda_G(m) \geqslant \min\left\{\frac{m}{2\sqrt{d}}, \frac{\sqrt{2Nm}}{d}\right\}$.*

Proof. Apply Theorem 2.4 to get a graph with $N' \geqslant N$, $N' \approx N$ vertices on the left, then remove an arbitrary subset of $N - N'$ vertices from the left hand side. This doesn't affect the expansion properties, but it destroys right-regularity. Apply Lemma 2.3 to correct this. □

2.2.3 Sum-product expanders

In this section, p will denote a prime, and $\mathbb{F}_p$ the finite field with p elements. The following result is implicit in [4, §4], and is based on a key "sum-product" lemma (Lemma 3.1) from [3], which is itself a statistical version of the sum-product theorems of Bourgain, Katz, and Tao [5], and Bourgain and Konyagin [6] for finite fields.

Proposition 2.7. *There exists an absolute constant $\xi_0 > 0$ such that for all primes p the following holds. Consider the bipartite graph $G_p = (\mathbb{F}_p^3, [4] \times \mathbb{F}_p, E)$ where a left vertex $(a, b, c) \in \mathbb{F}_p^3$ is adjacent to $(1, a)$, $(2, b)$, $(3, c)$, and $(4, a \cdot b + c)$ on the right. Then $\Lambda_{G_p}(m) \geqslant \min\left\{p^{0.9}, m^{1/3+\xi_0}\right\}$.*

Note that trivially $|\Gamma_{G_p}(S)| \geqslant |S|^{1/3}$, and the above states that not-too-large sets S expand by a sizeable amount more than the trivial bound. Using the above construction, we can now prove the following.

Theorem 2.8. *For all integers $N \geqslant 1$, there is an explicit construction of an $(N, n, 8, \Theta(N^{2/3}))$-right regular graph G which satisfies*

$$\Lambda_G(m) \geqslant \min\left\{\tfrac{1}{8}n^{0.9}, m^{1/3+\xi_0}\right\}.$$

(Here ξ_0 is the absolute constant from Proposition 2.7.)

Proof. Let p be the smallest prime such that $p^3 \geqslant N$; note that $N^{1/3} \leqslant p \leqslant 2N^{1/3}$. Construct the graph G_p, and a subgraph H of G_p by deleting an arbitrary $p^3 - N$ vertices on the left. Thus H has N vertices on left, $4p$ vertices on the right, is left-regular with degree 4 and satisfies, by Proposition 2.7, $\Lambda_H(m) \geqslant \min\left\{p^{0.9}, m^{1/3+\xi_0}\right\}$. Applying the transformation of Lemma 2.3 to H, we get an $(N, n, 8, d)$-right regular graph with $d = \lceil \frac{4N}{4p} \rceil \approx N^{2/3}$ and with the same expansion property. □

2.3 Distortion and spreading

For a vector $x \in \mathbb{R}^N$ and a subset $S \subseteq [N]$ of coordinates, we denote by $x_S \in \mathbb{R}^{|S|}$ the projection of x onto the coordinates in S. We abbreviate the complementary set of coordinates $[N] \setminus S$ to $\bar{S}$.

Definition 2.9 (Distortion of a subspace)**.** *For a subspace $X \subseteq \mathbb{R}^N$, we define*

$$\Delta(X) = \sup_{\substack{x \in X \\ x \neq 0}} \frac{\sqrt{N}\|x\|_2}{\|x\|_1}.$$

As we already noted in the introduction, instead of distortion it turns out to be more convenient to work with the following notion.

Definition 2.10. *A subspace $X \subseteq \mathbb{R}^N$ is* (t,ε)-spread *if for every $x \in X$ and every $S \subseteq [N]$ with $|S| \leqslant t$, we have*

$$\|x_{\bar{S}}\|_2 \geqslant \varepsilon \cdot \|x\|_2.$$

Let us begin with relating these two notions.

Lemma 2.11. *Suppose $X \subseteq \mathbb{R}^N$.*

a) *If X is (t,ε)-spread then*

$$\Delta(X) \leqslant \sqrt{\frac{N}{t}} \cdot \varepsilon^{-2};$$

b) *conversely, X is $\left(\frac{N}{2\Delta(X)^2},\ \frac{1}{4\Delta(X)}\right)$-spread.*

Proof. a). Fix $x \in X$; we need to prove that

$$\|x\|_1 \geqslant \sqrt{t}\varepsilon^2 \|x\|_2. \tag{2}$$

W.l.o.g. assume that $\|x\|_2 = 1$ and that $|x_1| \geqslant |x_2| \geqslant \ldots \geqslant |x_N|$. Applying Definition 2.10, we know that $\|x_{[t+1..N]}\|_2 \geqslant \varepsilon$. On the other hand, $\sum_{i=1}^t |x_i|^2 \leqslant 1$, therefore $|x_t| \leqslant \frac{1}{\sqrt{t}}$ and thus $\|x_{[t+1..N]}\|_\infty \leqslant \frac{1}{\sqrt{t}}$. And now we get (2) by the calculation

$$\|x\|_1 \geqslant \|x_{[t+1..N]}\|_1 \geqslant \frac{\|x_{[t+1..N]}\|_2^2}{\|x_{[t+1..N]}\|_\infty} \geqslant \sqrt{t}\varepsilon^2.$$

b). Let $t = \frac{N}{2\Delta(X)^2}$. Fix again $x \in X$ with $\|x\|_2 = 1$ and $S \subseteq [N]$ with $|S| \leqslant t$. By the bound on distortion, $\|x\|_1 \geqslant \frac{\sqrt{N}}{\Delta(X)}$. On the other hand,

$$\|x_S\|_1 \leqslant \sqrt{t} \cdot \|x_S\|_2 \leqslant \sqrt{t} = \frac{\sqrt{N/2}}{\Delta(X)},$$

hence $\|x_{\bar{S}}\|_1 = \|x\|_1 - \|x_S\|_1 \geqslant \frac{\sqrt{N}}{4\Delta(X)}$ and $\|x_{\bar{S}}\|_2 \geqslant \frac{\|x_{\bar{S}}\|_1}{\sqrt{N}} \geqslant \frac{1}{4\Delta(X)}$. □

Next, we note spreading properties of random subspaces (they will be needed only in the proof of Theorem 1.2). The following theorem is due to Kashin [23], with the optimal bound essentially obtained by Garnaev and Gluskin [15]. We note that such a theorem now follows from standard tools in asymptotic convex geometry, given the entropy bounds of Schütt [34] (see, e.g. Lemma B in [27]).

Theorem 2.12. *If A is a uniformly random $k \times N$ sign matrix, then with probability $1-o(1)$,*

$$\Delta(\ker(A)) \lesssim \sqrt{\frac{N}{k} \log\left(\frac{N}{k}\right)}.$$

Combining Theorem 2.12 with Lemma 2.11(b), we get:

Theorem 2.13. *If A is a uniformly random $k \times N$ sign matrix, then with probability $1-o(1)$, $\ker(A)$ is a $\left(\Omega\left(\frac{k}{\log(N/k)}\right), \Omega\left(\sqrt{\frac{k}{N\log(N/k)}}\right)\right)$-spread subspace.*

Finally, we introduce a "relative" version of Definition 2.10. It is somewhat less intuitive, but very convenient to work with.

Definition 2.14. *A subspace $X \subseteq \mathbb{R}^N$ is (t, T, ε)-*spread *($t \leqslant T$) if for every $x \in X$,*

$$\min_{\substack{S \subseteq [N] \\ |S| \leqslant T}} \|x_{\bar{S}}\|_2 \geqslant \varepsilon \cdot \min_{\substack{S \subseteq [N] \\ |S| \leqslant t}} \|x_{\bar{S}}\|_2.$$

Note that X is (t, ε)-spread if and only if it is $(0, t, \varepsilon)$-spread, if and only if it is $(1/2, t, \varepsilon)$-spread. (Note that t, T are not restricted to integers in our definitions.) One obvious advantage of Definition 2.14 is that it allows us to break the task of constructing well-spread subspaces into pieces.

Lemma 2.15. *Let $X_1, \ldots, X_r \subseteq \mathbb{R}^N$ be linear subspaces, and assume that X_i is $(t_{i-1}, t_i, \varepsilon_i)$-spread, where $t_0 \leqslant t_1 \leqslant \cdots \leqslant t_r$. Then $\bigcap_{i=1}^r X_i$ is $(t_0, t_r, \prod_{i=1}^r \varepsilon_i)$-spread.*

Proof. Obvious. □

3 An explicit weakly-spread subspace

Now our goal can be stated as finding an explicit construction that gets as close as possible to the probabilistic bound of Theorem 2.13. In this section we perform a (relatively simple) "initialization" step; the boosting argument (which is the most essential contribution of our paper) is deferred to Section 4. Below, for a matrix A, we denote by $\|A\|$ its operator norm, defined as $\sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$.

Lemma 3.1. *Let A be any $k \times d$ matrix whose columns $a_1, \ldots, a_d \in \mathbb{R}^k$ have ℓ_2-norm 1, and, moreover, for any $1 \leqslant i < j \leqslant d$, $|\langle a_i, a_j \rangle| \leqslant \tau$. Then $\ker(A)$ is $\left(\frac{1}{2\tau}, \frac{1}{2\|A\|}\right)$-spread.*

Proof. Fix $x \in \ker(A)$ and let $S \subseteq [d]$ be any subset with $t = |S| \leqslant \frac{1}{2\tau}$. Let A_S be the $k \times t$ matrix which arises by restricting A to the columns indexed by S, and let $\Phi = A_S^{\mathsf{T}} A_S$. Then Φ is the $t \times t$ matrix whose entries are $\langle a_i, a_j \rangle$ for $i, j \in S$, therefore we can write $\Phi = I + \Phi'$ where every entry of Φ' is bounded in magnitude by τ. It follows that all the eigenvalues of Φ lie in the range $[1 - t\tau, 1 + t\tau]$. We conclude, in particular, that $\|A_S\, y\|_2^2 \geqslant (1 - t\tau)\|y\|_2^2 \geqslant \frac{1}{2}\|y\|_2^2$ for every $y \in \mathbb{R}^t$.

Let $A_{\bar{S}}$ be the restriction of A to the columns in the complement of S. Since $x \in \ker(A)$, we have

$$0 = Ax = A_S x_S + A_{\bar{S}} x_{\bar{S}}$$

so that

$$\|A_{\bar{S}}x_{\bar{S}}\|_2 = \|A_S x_S\|_2 \geqslant \frac{1}{\sqrt{2}}\|x_S\|_2.$$

Since $\|A_{\bar{S}}x_{\bar{S}}\|_2 \leqslant \|A\| \cdot \|x_{\bar{S}}\|_2$, it follows that $\|x_S\|_2 \leqslant \sqrt{2}\|A\| \cdot \|x_{\bar{S}}\|_2$. Since $\|A\| \geqslant 1$, this implies $\|x_{\bar{S}}\|_2 \geqslant \frac{\|x\|_2}{2\|A\|}$. □

We now obtain matrices with small operator norm and near-orthogonal columns from explicit constructions of Kerdock codes.

Proposition 3.2. *For all positive integers d, k where k is a power of 4 satisfying $k \leqslant d \leqslant k^2/2$, there exists an explicit $k \times d$ matrix A with the following properties.*

1. *Every entry of A is either $\pm 1/\sqrt{k}$, and thus the columns $a_1, a_2, \ldots, a_d \in \mathbb{R}^k$ of A all have ℓ_2-norm 1,*
2. *For all $1 \leqslant i < j \leqslant d$, $|\langle a_i, a_j\rangle| \leqslant 1/\sqrt{k}$, and*
3. *$\|A\| \leqslant \sqrt{\lceil \frac{d}{k} \rceil}$.*

Proof. The proof is based on a construction of mutually unbiased bases over the reals using Kerdock codes [25, 7]. First, let us recall that for k a power of 2, the Hadamard code of length k is a subspace of $\mathbb{F}_2^k$ of size k containing the k linear functions $L_a : \mathbb{F}_2^{\log_2 k} \to \mathbb{F}_2$, where for $a, x \in \mathbb{F}_2^{\log_2 k}$, $L_a(x) = a \cdot x$ (computed over $\mathbb{F}_2$). A Kerdock code is the union of a Hadamard code $H \subseteq \mathbb{F}_2^k$ and a collection of its cosets $\{f + H \mid f \in \mathcal{F}\}$, where $\mathcal{F}$ is a set of quadratic bent functions with the property that for all $f \neq g \in \mathcal{F}$, the function $f + g$ is also bent.[2]

When k is a power of 4, it is known (see [25] and also [29, Chap. 15, Sec. 5]) that one can construct an explicit set $\mathcal{F}$ of $(\frac{k}{2} - 1)$ such bent functions. (A simpler construction of $(\sqrt{k} - 1)$ such quadratic functions appears in [7].) The cosets of these functions together with the Hadamard code (the trivial coset) give an explicit Kerdock code of length k that has $k^2/2$ codewords. Interpreting binary vectors of length k as unit vectors with $\pm 1/\sqrt{k}$ entries, every coset of the Hadamard code gives an orthonormal basis of $\mathbb{R}^k$. The $k/2$ cosets comprising the Kerdock code thus yield $k/2$ orthonormal bases $B_1, B_2, \ldots, B_{k/2}$ of $\mathbb{R}^k$ with the property that for every pair $\{v, w\}$ of vectors in different bases, one has $|\langle v, w\rangle| = 1/\sqrt{k}$. (Such bases are called mutually unbiased bases.)

For any d, $k \leqslant d \leqslant k^2/2$, write $d = qk + r$ where $0 \leqslant r < k$. We construct our $k \times d$ matrix A to consist of $[B_1 \ldots B_q]$ followed by, in the case of $r > 0$, any r columns of B_{q+1}. The first two properties of A are immediate from the property of the bases B_i. To bound the operator norm, note that being an orthonormal basis, $\|B_i\| = 1$ for each i. A simple application of Cauchy-Schwartz then shows that $\|A\| \leqslant \sqrt{\lceil d/k \rceil}$. □

Plugging in the matrices guaranteed by Proposition 3.2 into Lemma 3.1, we can conclude the following.

[2] A function $f : \mathbb{F}_2^a \to \mathbb{F}_2$ for a even is said to be *bent* if it is maximally far from all linear functions, or equivalently if all its Fourier coefficients have absolute value $1/2^{a/2}$.

Theorem 3.3. *For every integer k that is a power of 4 and every integer d such that*

$$k \leqslant d \leqslant k^2/2, \tag{3}$$

there exists a $\left(\frac{\sqrt{k}}{2}, \frac{1}{4}\sqrt{\frac{k}{d}}\right)$-spread subspace $L \subseteq \mathbb{R}^d$ with $\mathsf{codim}(L) \leqslant k$, specified as the kernel of an explicit $k \times d$ sign matrix.

These subspaces will be used as "inner" subspaces in an expander-based construction (Theorem 4.3) to get a subspace with even better spreading properties.

4 Boosting spreading properties via expanders

4.1 The Tanner construction

Definition 4.1 (Subspaces from bipartite graphs)**.** *Given a bipartite graph $G = (\{1, 2, \ldots, N\}, V_R, E)$ such that every vertex in V_R has degree d, and a subspace $L \subseteq \mathbb{R}^d$, we define the subspace $X = X(G, L) \subseteq \mathbb{R}^N$ by*

$$X(G, L) = \{x \in \mathbb{R}^N \mid x_{\Gamma_G(j)} \in L \text{ for every } j \in V_R\} \ . \tag{4}$$

The following claim is straightforward.

Claim 1. *If $n = |V_R|$, then $\mathsf{codim}(X(G, L)) \leqslant \mathsf{codim}(L)n$, that is $\mathsf{dim}(X(G, L)) \geqslant N - (d - \mathsf{dim}(L))n$.*

Remark 1 (Tanner's code construction)**.** Our construction is a continuous analog of Tanner's construction of error-correcting codes [37]. Tanner constructed codes by identifying the vertices on one side of a bipartite graph with the bits of the code and identifying the other side with constraints. He analyzed the performance of such codes by examining the girth of the bipartite graph. Sipser and Spielman [35] showed that graph expansion plays a key role in the quality of such codes, and gave a linear time decoding algorithm to correct a constant fraction of errors. In the coding world, the special case when L is the $(d-1)$-dimensional subspace $\{y \in \mathbb{R}^d \mid \sum_{\ell=1}^{d} y_\ell = 0\}$ corresponds to the low-density parity check codes of Gallager [14]. In this case, the subspace is specified as the kernel of the bipartite adjacency matrix of G.

4.2 The spread-boosting theorem

We now show how to improve spreading properties using the above construction.

Theorem 4.2. *Let G be an (N, n, D, d)-graph with expansion profile $\Lambda_G(\cdot)$, and let $L \subseteq \mathbb{R}^d$ be a (t, ε)-spread subspace. Then for every T_0, $0 < T_0 \leqslant N$, $X(G, L)$ is $\left(T_0, \frac{t}{D}\Lambda_G(T_0), \frac{\varepsilon}{\sqrt{2D}}\right)$-spread.*

Proof. Fix $x \in X(G,L)$ with $\|x\|_2 = 1$. Fix also $S \subseteq [N]$ with $|S| \leqslant T$, where $T = \frac{t}{D}\Lambda_G(T_0)$. We then need to prove that

$$\|x_{\bar{S}}\|_2 \geqslant \frac{\varepsilon}{\sqrt{2D}} \min_{|B| \leqslant T_0} \|x_{\bar{B}}\|_2. \tag{5}$$

Let

$$Q = \{j \in [n] : |\Gamma(j) \cap S| > t\},$$

and

$$B = \{i \in S : \Gamma(i) \subseteq Q\}.$$

Then

$$t|Q| < E(S, \Gamma(S)) \leqslant D|S| \leqslant DT,$$

therefore

$$|Q| < \frac{DT}{t} = \Lambda_G(T_0)\ .$$

On the other hand, we have $|Q| \geqslant |\Gamma(B)|$, and hence $|\Gamma(B)| < \Lambda_G(T_0)$. By the definition of the expansion profile, this implies that $|B| < T_0$, and therefore (see (5)) we are only left to show that

$$\|x_{\bar{S}}\|_2 \geqslant \frac{\varepsilon}{\sqrt{2D}} \cdot \|x_{\bar{B}}\|_2 \tag{6}$$

for our particular B.

Note first that

$$\|x_{\bar{B}}\|_2^2 = \|x_{\bar{S}}\|_2^2 + \|x_{S\setminus B}\|_2^2. \tag{7}$$

Next, since every vertex in $S \setminus B$ has at least one neighbor in $\Gamma(S) \setminus Q$, we have

$$\sum_{j \in \Gamma(S)\setminus Q} \|x_{\Gamma(j)}\|_2^2 \geqslant \|x_{S\setminus B}\|_2^2. \tag{8}$$

Since $x \in X(G,L)$, L is (t,ε)-spread, and $|\Gamma(j) \cap S| \leqslant t$ for any $j \in \Gamma(S) \setminus Q$,

$$\sum_{j \in \Gamma(S)\setminus Q} \|x_{\Gamma(j)\setminus S}\|_2^2 \geqslant \varepsilon^2 \cdot \sum_{j \in \Gamma(S)\setminus Q} \|x_{\Gamma(j)}\|_2^2. \tag{9}$$

Finally,

$$\sum_{j \in \Gamma(S)\setminus Q} \|x_{\Gamma(j)\setminus S}\|_2^2 \leqslant \sum_{j \in [n]} \|x_{\Gamma(j)\setminus S}\|_2^2 \leqslant D \cdot \|x_{\bar{S}}\|_2^2. \tag{10}$$

(7)-(10) imply

$$\|x_{\bar{S}}\|_2^2 \geqslant \frac{\varepsilon^2}{D}(\|x_{\bar{B}}\|_2^2 - \|x_{\bar{S}}\|_2^2).$$

Since $\varepsilon \leqslant 1$ and $D \geqslant 1$, (6) (and hence Theorem 4.2) follows. □

4.3 Putting things together

In this section we assemble the proofs of Theorems 1.1 and 1.2 from the already available blocks (which are Theorems 2.8, 2.6, 2.13, 3.3 and 4.2). Let us first see what we can do using expanders from Theorem 2.8.

4.3.1 First step: Boosting with sum-product expanders

The main difference between the explicit construction of Theorem 3.3 and the probabilistic result (Theorem 2.13) is the order of magnitude of t (the parameter from Definition 2.10). As we will see in the next section, this difference is very principal, and our first goal is to *somewhat* close the gap with an *explicit* construction.

Theorem 4.3. *Fix an arbitrary constant $\beta_0 < \min\left\{0.08, \frac{3}{8}\xi_0\right\}$, where ξ_0 is the constant from Theorem 2.8. Then for all sufficiently large $N \in \mathbb{N}$ and $\eta \geqslant N^{-2\beta_0/3}$ there exists an explicit subspace $X \subseteq \mathbb{R}^N$ with $\mathsf{codim}(X) \leqslant \eta N$ which is $(N^{\frac{1}{2}+\beta_0}, \eta^{O(1)})$-spread.*

Proof. In everything that follows, we assume that N is sufficiently large. The desired X will be of the form $X(G, L)$, where G is supplied by Theorem 2.8, and L by Theorem 3.3. More specifically, let G be the explicit $(N, n, 8, d)$-right regular graph from Theorem 2.8 with $d \approx N^{2/3}$ (and hence $n \approx N^{1/3}$). Using Theorem 2.8, one can check that for $m \leqslant N^{\frac{1}{2}+\beta_0}$, we have

$$\Lambda_G(m) \geqslant md^{\beta_0 - \frac{1}{2}} \ . \tag{11}$$

Indeed, since $n \approx N^{1/3}$ and $d \approx N^{2/3}$, the inequality $\frac{1}{8}n^{0.9} \geqslant md^{\beta_0-\frac{1}{2}}$ follows (for large N) from $\beta_0 < 0.08$, and the inequality $m^{\frac{1}{3}+\xi_0} \geqslant md^{\beta_0-\frac{1}{2}}$ follows from $\beta_0 < \frac{3}{8}\xi_0$.

By our assumption $\eta \geqslant N^{-2\beta_0/3} \geqslant N^{-0.1}$, along with $d \approx N^{2/3}$, we observe that $d \leqslant o\left(\eta d\right)^2$. Hence (cf. the statement of Theorem 3.3), we can find $k \leqslant \frac{\eta d}{8}$, $k \approx \eta d$ that is a power of 4 and also satisfies the restrictions (3). Let L be an explicit $\left(\Omega\left(\sqrt{\eta d}\right), \Omega\left(\sqrt{\eta}\right)\right)$-spread subspace guaranteed by Theorem 3.3.

The bound on codimension of $X(G, L)$ is obvious: $\mathsf{codim}(X(G, L)) \leqslant kn \leqslant \frac{\eta d n}{8} \leqslant \eta N$.

For analyzing spreading properties of $X(G, L)$, we observe that $\eta \geqslant N^{-2\beta_0/3}$ implies $\eta d \gtrsim d^{1-\beta_0}$, hence L is $(\Omega(d^{\frac{1}{2}-\frac{\beta_0}{2}}), \eta^{O(1)})$-spread. By Theorem 4.2 and (11), for every $T \leqslant N^{\frac{1}{2}+\beta_0}$, we know that $X(G, L)$ is $(T, \Omega(d^{\frac{\beta_0}{2}})T, \eta^{O(1)})$-spread In particular, for such T, $X(G, L)$ is $(T, N^{\Omega(1)}T, \eta^{O(1)})$-spread.

Applying Lemma 2.15 with the same spaces $X_1 := \cdots := X_r := X(G, L)$ and suitably large constant $r \approx 1/\beta_0 = O(1)$, we conclude that $X(G, L)$ is $\left(\frac{1}{2}, N^{\frac{1}{2}+\beta_0}, \eta^{O(1)}\right)$-spread, completing the proof. □

4.3.2 Second step: Handling large sets based on spectral expanders

The sum-product expanders of Theorem 2.8 behave poorly for very large sets (i.e., as $m \to N$, the lower bound on $\Lambda_G(m)$ becomes constant from some point). The spectral expanders of Theorem 2.6 behave poorly for small sets, but their expansion still improves as $m \to N$.

In this section, we finish the proofs of Theorems 1.1 and 1.2 by exploring strong sides of both constructions. We begin with Theorem 1.2 as it is conceptually simpler (we need only spectral expanders, do not rely on Theorem 4.3, and still use only one fixed space $X(G, L)$).

Proof of Theorem 1.2. By Theorem 2.6 there exists an explicit $(N, n, 4, d)$-right regular graph G with

$$N^{\Omega\left(\frac{1}{\log\log N}\right)} \leqslant d \leqslant N^{\frac{1}{2\log\log N}} \tag{12}$$

which has $\Lambda_G(m) \geqslant \min\left\{\frac{m}{2\sqrt{d}}, \frac{\sqrt{2Nm}}{d}\right\}$. Let $k = \lfloor \frac{\eta}{4} d \rfloor$; our desired (probabilistic) space is then $X(G, \ker(A))$, where A is a uniformly random $k \times d$ sign matrix (due to the upper bound in (12), this uses at most $d^2 \leqslant N^{\frac{1}{\log\log N}}$ random bits). Recalling that $\eta > 0$ is an absolute constant, by Theorem 2.13 $\ker(A)$ is an $(\Omega(d), \Omega(1))$-spread subspace almost surely.

The bound on codimension is again simple: $\mathsf{codim}(X(G, \ker(A))) \leqslant kn \leqslant \eta N$.

For analyzing spreading properties of X, let $m_0 = 8N/d$ (which is the "critical" point where $\frac{m_0}{2\sqrt{d}} = \frac{\sqrt{2Nm_0}}{d}$.) Then Theorem 4.2 says that $X(G, L)$ is

a. $\Big(T, \Omega(\sqrt{d})T, \Omega(1)\Big)$-spread subspace for $T \leqslant m_0$, and

b. $\Big(T, \Omega(\sqrt{NT}), \Omega(1)\Big)$-spread subspace for $m_0 \leqslant T \leqslant N$.

And now we are once more applying Lemma 2.15 with $X_1 := X_2 := \ldots := X_r := X(G, L)$. In $O(\log_d m_0) = O(\log\log N)$ applications of (a) with $T \leqslant m_0$, we conclude that $X(G, L)$ is $(\frac{1}{2}, m_0, (\log N)^{-O(1)})$-spread. In $O(\log\log N)$ additional applications of (b) with $T \geqslant m_0$, we conclude that $X(G, L)$ is $(\frac{1}{2}, \Omega(N), (\log N)^{-O(1)})$-spread.

Since $X(G, L)$ is an $(\Omega(N), (\log N)^{-O(1)})$-spread subspace, the statement of Theorem 1.2 immediately follows from Lemma 2.11(a).

Proof of Theorem 1.1. This is our most sophisticated construction: we use a series of $X(G, L)$ for *different* graphs G, and the "inner" spaces L will come from Theorem 4.3. In what follows, we assume that N is sufficiently large (obviously for $N = O(1)$, every non-trivial subspace has bounded distortion).

To get started, let us denote

$$\widetilde{\eta} = \frac{\eta}{(\log\log N)^2},$$

and let us first construct and analyze subspaces $X(G, L)$ needed for our purposes individually. For that purpose, fix (for the time being) any value of m with

$$1 \leqslant m \leqslant \delta \widetilde{\eta}^{2\beta_0/3} N, \tag{13}$$

δ a sufficiently small constant and β_0 is the constant from Theorem 4.3.

Applying Theorem 2.6 (with $d := N/m$), we get, for some $d = \Theta(N/m)$, an explicit $(N, n, 4, d)$-right regular graph G_m with $\Lambda_{G_m}(m) \geqslant \Omega(d^{-1/2})m$. Note that (13) implies $\widetilde{\eta} \geqslant d^{-2\beta_0/3}$ (provided the constant δ is small enough), and thus all conditions of Theorem 4.3

with $N := d,\ \eta := \widetilde{\eta}$ are met. Applying that theorem, let $L_m \subseteq \mathbb{R}^d$ be an explicit subspace with $\mathsf{codim}(L_m) \leqslant \widetilde{\eta} d$ that is a $(d^{\frac{1}{2}+\beta_0}, (\eta/\log\log N)^{O(1)})$-spread subspace. Consider the space $X(G_m, L_m) \subseteq \mathbb{R}^N$.

Since $D = 4$ is a constant, we have

$$\mathsf{codim}(X(G_m, L_m)) \lesssim \widetilde{\eta} N = \frac{\eta N}{(\log\log N)^2}.$$

And Theorem 4.2 (applied to $T := m$) implies (recalling $\Lambda_{G_m}(m) \gtrsim d^{-1/2}m$, $t = d^{\frac{1}{2}+\beta_0}$, $d = \Theta(N/m)$) that $X(G_m, L_m)$ is a $\left(m,\ \Omega\left(\left(\frac{N}{m}\right)^{\beta_0}\right) m,\ (\eta/\log\log N)^{O(1)}\right)$-spread subspace. We note that it is here that we crucially use the fact that L_m has spreading properties for $t \gg d^{1/2}$ (t is the parameter from Definition 2.10) so that we more than compensate for the factor $\sqrt{d}$ loss in Theorem 1.2 caused by the relatively poor expansion rate of spectral expanders.

We will again apply Lemma 2.15, but the spaces X_i will now be distinct. In particular, for $i \in \mathbb{N}$ define $X_i = X(G_{t_i}, L_{t_i})$, where

$$t_i = N \cdot \left(\frac{\varepsilon}{N}\right)^{(1-\beta_0)^i},$$

for some sufficiently small constant ε, $0 < \varepsilon < 1$. It is easy to see that for some $r = O(\log\log N)$, we have $t_r \leqslant \delta\tilde{\eta}^{2\beta_0/3} N$ and $t_r \gtrsim \left(\delta\tilde{\eta}^{2\beta_0/3}\right)^2 N$.

Then for $X = \bigcap_{i=0}^{r-1} X_i$ we have $\mathsf{codim}(X) \lesssim r\frac{\eta N}{(\log\log N)^2} \lesssim \frac{\eta N}{\log\log N}$. In particular, $\mathsf{codim}(X) \leqslant \eta N$ for sufficiently large N.

By the above argument based on Theorem 4.2 and the choice of the t_i's, it is easily seen that X_i is a $(t_i, t_{i+1}, (\eta/(\log\log N))^{O(1)})$-spread subspace. By Lemma 2.15, X is a $(\varepsilon, t_r, (\eta/(\log\log N))^{O(\log\log N)})$-spread subspace, or equivalently a $(t_r, (\eta/(\log\log N))^{O(\log\log N)})$-spread subspace. Since we also have $t_r \geqslant (\eta/(\log\log N))^{O(1)} N$, the required bound on $\Delta(X)$ follows from Lemma 2.11(a).

5 Discussion

We have presented explicit subspaces $X \subseteq \mathbb{R}^N$ of dimension $(1-\eta)N$ with distortion $(\eta^{-1}\log\log N)^{O(\log\log N)}$ and, using $N^{o(1)}$ random bits, distortion $\eta^{-O(\log\log N)}$. It is natural to wonder whether better explicit constructions of expanders can give rise to better bounds. We make some remarks about this possibility.

1. **The GUV and CRVW expander families.** The next two theorems essentially follow from [17] and [8], respectively (after an appropriate application of Lemma 2.3).

   Theorem 5.1 ([17])**.** *For each fixed $0 < c, \varepsilon \leqslant 1$, and for all integers N, K with $K \leqslant N$, there is an explicit construction of an (N, n, D, d)-right regular graph G with $D \lesssim ((\log N)/\varepsilon)^{2+2/c}$ and $d \geqslant N/(DK^{1+c})$ and such that $\Lambda_G(m) \geqslant (1-\varepsilon)D \cdot \min\{K, m\}$.*

Theorem 5.2 ([8]). *For every fixed $0 < \varepsilon < 1$ and all sufficiently large values N and d there exist $n \leqslant N$, $D \leqslant 2^{O((\varepsilon^{-1}\log\log d)^3)}$ and an explicit (N, n, D, d)-right regular bipartite graph G with $\Lambda_G(m) \geqslant (1-\varepsilon)D \cdot \min\{\Omega(N/d), m\}$.*

The main problem for us in both these constructions is that D must grow with N and d, respectively. By plugging in the explicit subspaces of Theorem 3.3 into Theorem 4.2 with the GUV-expanders from Theorem 5.1, one can achieve distortions $\Delta(X) \approx \exp(\sqrt{\log N \log\log N})$ for $X \subseteq \mathbb{R}^N$ with $\mathsf{dim}(X) \geqslant N/2$. Using the GUV-expanders (in place of the sum-product expanders) together with spectral expanders in a construction similar to the proof of Theorem 1.1 would yield a distortion bound of $(\log N)^{O(\log\log N)}$.

2. **Very good expansion for large sets.** If it were possible to construct an (N, n, D, d)-right regular bipartite graph H with $D = O(1)$ and such that for every $S \subseteq V_L$ with $|S| \geqslant N^{1-\beta}$, we had $|\Gamma(S)| = \Omega(n)$, then we would be able to achieve $O(1)$ distortion using only $O(d^2 + N^\delta)$ random bits for any $\delta > 0$ (in fact, we could use only $O(d + N^\delta)$ random bits with [27]).

   The idea would be to follow the proof of Theorem 1.2, but only for $O(1)$ steps to show the subspace is $(N^{1-\beta}, \Omega(1))$-spread. Then we would intersect this with a subspace $X(H, L)$, where $L \subseteq \mathbb{R}^d$, with the latter subspace generated as the kernel of a random sign matrix (requiring d^2 bits). Unfortunately, [31, Th. 1.5] shows that in order to achieve the required expansion property, one has to take $D \geqslant \Omega(\beta \log N)$.

Acknowledgments

We are grateful to Avi Wigderson for several enlightening discussions, and especially his suggestion that the sum-product expanders of [3, 4] should be relevant. Using the sum-product expanders in place of GUV-expanders in Section 4.3.1, we were able to improve our distortion bound from $(\log N)^{O(\log\log N)}$ to $(\log N)^{O(\log\log\log N)}$. We are also thankful to an anonymous referee for several useful remarks.

References

[1] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks. In *Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)*, volume 72, pages 15–19, 1988.

[2] S. Artstein-Avidan and V. D. Milman. Logarithmic reduction of the level of randomness in some probabilistic geometric constructions. *J. Funct. Anal.*, 235(1):297–329, 2006.

[3] B. Barak, R. Impagliazzo, and A. Wigderson. Extracting randomness using few independent sources. *SIAM Journal on Computing*, 36(4):1095–1118, 2006.

[4] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov, and A. Wigderson. Simulating independence: New constructions of condensers, Ramsey graphs, dispersers, and extractors. In *Proceedings of the 46th ACM Symposium on Theory of Computing*, pages 1–10, 2005.

[5] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. *Geom. Funct. Anal.*, 14(1):27–57, 2004.

[6] J. Bourgain and S. V. Konyagin. Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order. *C. R. Math. Acad. Sci. Paris*, 337(2):75–80, 2003.

[7] P. J. Cameron and J. J. Seidel. Quadratic forms over GF(2). *Indag. Math.*, 35:1–8, 1973.

[8] M. R. Capalbo, O. Reingold, S. P. Vadhan, and A. Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, pages 659–668, 2002.

[9] R. A. DeVore. Deterministic constructions of compressed sensing matrices. Manuscript, 2007.

[10] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006.

[11] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49(3):906–931, 1989.

[12] P. Erdös. A theorem of Sylvester and Schur. *J. London Math. Soc.*, 9:282–288, 1934.

[13] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.

[14] R. G. Gallager. *Low-Density Parity-Check Codes*. MIT Press, 1963.

[15] A. Garnaev and E. D. Gluskin. The widths of Euclidean balls. *Doklady An. SSSR.*, 277:1048–1052, 1984.

[16] V. Guruswami, J. Lee, and A. Wigderson. Euclidean sections of with sublinear randomness and error-correction over the reals. In *12th International Wrokshop on Randomization and Combinatorial Optimization: Algorithms and Techniques (RANDOM)*, pages 444–454, 2008.

[17] V. Guruswami, C. Umans, and S. P. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. In *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity*, pages 96–108, 2007.

[18] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.

[19] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM*, 53(3):307–323, 2006.

[20] P. Indyk. Uncertainty principles, extractors, and explicit embeddings of L_1 into L_2. In *Proceedings of the 39th Annual ACM Symposium on the Theory of Computing*, pages 615–620, 2007.

[21] P. Indyk. Explicit constructions for compressed sensing of sparse signals. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 30–33, 2008.

[22] W. B. Johnson and G. Schechtman. Finite dimensional subspaces of L_p. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 837–870. North-Holland, Amsterdam, 2001.

[23] B. S. Kashin. The widths of certain finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, 41(2):334–351, 478, 1977.

[24] B. S. Kashin and V. N. Temlyakov. A remark on compressed sensing. Available at `http://www.dsp.ece.rice.edu/cs/KT2007.pdf`, 2007.

[25] A. M. Kerdock. A class of low-rate nonlinear binary codes. *Inform. Control*, 20:182–187, 1972.

[26] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[27] S. Lovett and S. Sodin. Almost Euclidean sections of the N-dimensional cross-polytope using $O(N)$ random bits. Electronic Colloquium on Computational Complexity, Report TR07-012, 2007.

[28] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[29] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland, 1977.

[30] V. Milman. Topics in asymptotic geometric analysis. *Geom. Funct. Anal.*, (Special Volume, Part II):792–815, 2000. GAFA 2000 (Tel Aviv, 1999).

[31] J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM J. Discrete Math.*, 13(1):2–24 (electronic), 2000.

[32] W. Rudin. Trigonometric series with gaps. *J. Math. Mech.*, 9:203–227, 1960.

[33] G. Schechtman. Random embeddings of Euclidean spaces in sequence spaces. *Israel J. Math.*, 40(2):187–192, 1981.

[34] C. Schütt. Entropy numbers of diagonal operators between symmetric Banach spaces. *J. Approx. Theory*, 40(2):121–128, 1984.

[35] M. Sipser and D. A. Spielman. Expander codes. *IEEE Trans. Inform. Theory*, 42(6, part 1):1710–1722, 1996. Codes and complexity.

[36] S. Szarek. Convexity, complexity, and high dimensions. In *International Congress of Mathematicians. Vol. II*, pages 1599–1621. Eur. Math. Soc., Zürich, 2006.

[37] R. M. Tanner. A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5):533–547, 1981.

[38] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

[39] G. Zémor. On expander codes. *IEEE Transactions on Information Theory*, 47(2):835–837, 2001.