Euclidean sections of ℓ_1^N with sublinear randomness and error-correction over the reals

Venkatesan Guruswami*, James R. Lee**, and Avi Wigderson

[1] Department of Comp. Sci. & Eng., University of Washington, and (on leave at) School of Mathematics, Institute for Advanced Study, Princeton
[2] Department of Comp. Sci. & Eng., University of Washington
[3] School of Mathematics, Institute for Advanced Study, Princeton

Abstract. It is well-known that $\mathbb{R}^N$ has subspaces of dimension proportional to N on which the ℓ_1 and ℓ_2 norms are uniformly equivalent, but it is unknown how to construct them explicitly. We show that, for any $\delta > 0$, such a subspace can be generated using only N^δ random bits. This improves over previous constructions of Artstein-Avidan and Milman, and of Lovett and Sodin, which require $O(N \log N)$, and $O(N)$ random bits, respectively.

Such subspaces are known to also yield error-correcting codes over the reals and compressed sensing matrices. Our subspaces are defined by the kernel of a relatively sparse matrix (with at most N^δ non-zero entries per row), and thus enable compressed sensing in near-linear $O(N^{1+\delta})$ time. As in the work of Guruswami, Lee, and Razborov, our construction is the continuous analog of a Tanner code, and makes use of expander graphs to impose a collection of local linear constraints on vectors in the subspace. Our analysis is able to achieve *uniform* equivalence of the ℓ_1 and ℓ_2 norms (independent of the dimension). It has parallels to iterative decoding of Tanner codes, and leads to an analogous near-linear time algorithm for error-correction over reals.

1 Introduction

Given $x \in \mathbb{R}^N$, one has the straightforward inequality $\|x\|_2 \leq \|x\|_1 \leq \sqrt{N}\|x\|_2$. Classical results of Figiel, Lindenstrauss, and Milman [FLM77] and Kasin [Kas77] show that for every $\eta > 0$, there exists a constant $C(\eta)$ and a subspace $X \subseteq \mathbb{R}^N$ with $\mathsf{dim}(X) \geq (1-\eta)N$ such that for every $x \in X$,

$$C(\eta)\sqrt{N}\|x\|_2 \leq \|x\|_1 \leq \sqrt{N}\|x\|_2.$$

We say that such a subspace has distortion at most $C(\eta)$, where for a subspace $X \subseteq \mathbb{R}^N$, we define the *distortion of* X as the quantity

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

* Supported by a Packard Fellowship and NSF grant CCR-0324906 to the IAS.
** Research supported by NSF CAREER award CCF-0644037.

The distortion always lies in the range $1 \leqslant \Delta(X) \leqslant \sqrt{N}$, and describes the extent to which the ℓ_2 mass of vectors in X is spread among the coordinates.

It is known that subspaces with good distortion give rise to error-correcting codes over the reals and compressed sensing matrices [KT07,Don06]. When viewed as embeddings of ℓ_2^n into ℓ_1^N (hence the terminology "distortion"), they are useful for problems like high-dimensional nearest-neighbor search [Ind06]. We discuss connections of our work to coding over reals in Section 3.

The existence proofs of [FLM77,Kas77] proceed by showing that a random subspace (for various notions of "random") satisfies the above conditions with positive probability. The problem of *explicit* constructions of these subspaces has been raised by a number of authors; see, e.g. [Sza06, Sec. 4], [Mil00, Prob. 8], [JS01, Sec. 2.2].

Although no explicit construction is known, there has been progress on reducing the amount of randomness needed to construct such subspaces. Kasin's proof [Kas77] is particularly amenable to such analysis because the subspaces he constructs are kernels of uniformly random sign matrices. This immediately gives rise to an algorithm which produces such a matrix using $O(N^2)$ random bits. Previous approaches to partial derandomization construct sign matrices whose entries are non-independent random signs; indeed, Artstein-Avidan and Milman reduce the randomness requirement to $O(N \log N)$ using random walks on expander graphs [AAM06], and Lovett and Sodin [LS07] improve this to $O(N)$ random bits by employing, in addition, families of $\Theta(\log N)$-wise independent random variables. We remark that the pseudorandom generator approach of Indyk [Ind06] can be used to efficiently construct such subspaces using $O(N \log^2 N)$ random bits.

As pointed out in [LS07], since these direct approaches require taking a union bound over an (exponentially large) ε-net, it is unlikely that they can be pushed beyond a linear dependence on the number of random bits. In contrast, the approaches of Indyk [Ind07] and Guruswami, Lee, and Razborov [GLR08] are inspired by work in randomness extraction and the theory of error-correcting codes. The latter paper uses a continuous variant of *expander codes* to deterministically construct a subspace $X \subseteq \mathbb{R}^N$ satisfying $\mathsf{dim}(X) \geq (1 - o(1))N$ and,

$$\Delta(X) \leqslant (\log N)^{O(\log \log \log N)} .$$

Even using sub-linear randomness, they are only able to achieve a distortion of $\mathrm{poly}(\log N)$. In the present paper, we continue with the expander codes approach of [GLR08], with a different construction and analysis. As a result, we are able to produce almost-Euclidean sections of ℓ_1^N with constant distortion and proportional dimension, while using only N^δ random bits for any $\delta > 0$. In Section 2 and Remark 1, we discuss how our analysis overcomes some difficulties from [GLR08].

In the next section, we show that given a subspace $X \subseteq \mathbb{R}^n$ with $\mathsf{dim}(X) \geq (1-\eta)n$, for every $N \geq n$ there exists a simple, explicit construction of a subspace $X' \subseteq \mathbb{R}^N$ satisfying $\mathsf{dim}(X') \geq (1 - 2\eta)N$ and $\Delta(X') \leq N^{O(\frac{1}{\log n})}\Delta(X)$. If

X is the kernel of a sign matrix, then so is X'. By choosing $n = N^{\delta/2}$ and generating X as the kernel of a random sign matrix, we achieve a construction with $\Delta(X') = 2^{O(1/\delta)}$ distortion using at most N^δ random bits.

1.1 Preliminaries

We use $[M]$ to denote the set $\{1, 2, \ldots, M\}$. For $x \in \mathbb{R}^N$ and a subset $I \subseteq [N]$, we denote by $x_I \in \mathbb{R}^{|I|}$ the restriction of x to the coordinates in I.

Definition 1 (Well spread subspaces) *A subspace $L \subseteq \mathbb{R}^m$ is said to be (b, ρ)-spread if for every $y \in L$, and every set $S \subseteq [m]$ size at least $m - b$, $\|y_S\|_2 \geqslant \rho\|y\|_2$.*

As stated below, there is a straightforward relation between spread subspaces and distortion (see, e.g. [GLR08, Lemma 2.11]), but the former notion is a well-suited to our arguments

Lemma 1. *Suppose $X \subseteq \mathbb{R}^N$.*

1. *If X is (b, ρ)-spread then*
$$\Delta(X) \leq \sqrt{\frac{N}{b}} \cdot \rho^{-2};$$
2. *Conversely, X is $\left(\frac{N}{2\Delta(X)^2}, \frac{1}{4\Delta(X)}\right)$-spread.*

We will make use of the following (non-constructive) result on the existence of well-spread subspaces; it is due to Kasin [Kas77], with the optimal bound obtained by Garnaev and Gluskin for uniformly random subspaces [GG84]. The proof that sign matrices suffice is now standard, given the covering estimates of Schütt [Sch84]; see e.g. [LS07, Lemma B].

Theorem 2. *For all integers $1 \leqslant k < d$, there exists a subspace $Y \subseteq \mathbb{R}^d$ of dimension at least $d - k$, specified as the kernel of a $k \times d$ sign matrix, such that $\Delta(Y) \leqslant O\left(\sqrt{\frac{d}{k}\log\frac{d}{k}}\right)$, and so by Lemma 1, Y is*
$$\left(\Omega\left(\frac{k}{\log(d/k)}\right), \Omega\left(\sqrt{\frac{k}{d\log(d/k)}}\right)\right)\text{-spread.}$$
In fact, a random such matrix has this property with probability $1 - o_d(1)$.

Definition 3 (Subspaces from regular graphs) *Given an undirected d-regular graph $G = (V, E)$ with N edges and a subspace $L \subseteq \mathbb{R}^d$, we define the subspace $T(G, L) \subseteq \mathbb{R}^N$ by*
$$T(G, L) = \{x \in \mathbb{R}^N \mid x_{\Gamma(v)} \in L \text{ for every } v \in V\}\ . \tag{1}$$
where $\Gamma(v)$ is the set of d edges incident on v in some fixed order.

The definition of the subspace $T(G, L)$ is inspired by the construction of *expander codes,* following Sipser and Spielman [SS96] and Tanner [Tan81] in the case of finite fields, and its adaptation to the reals by Guruswami, Lee, and Razborov [GLR08].

Definition 4 (Expander) *A simple, undirected graph G is said to be an (n, d, λ)-expander if G has n vertices, is d-regular, and the second largest eigenvalue of the adjacency matrix of G in absolute value is at most λ.*

For a graph $G = (V, E)$ (which will be clear from context) and $W \subseteq V$, we denote by $E(W)$ the set of edges both of whose endpoints lie in W. For two subsets $X, Y \subseteq V$ (which could intersect), we denote by $E(X, Y)$ the (multi)set of edges with one endpoint in X and the other in Y. Recall that for a vertex $v \in V$, $\Gamma(v) \subseteq E$ is the set of edges incident upon v.

2 Derandomized sections

Following [GLR08], we now show that if L is well-spread and G is an expander graph, then $T(G, L)$ is itself well-spread. This immediately implies the ability to create large dimensional low-distortion subspaces from those with smaller dimension. In Remark 1, we discuss how our analysis is able to overcome the apparent barrier in [GLR08]. Finally, in Section 2.2, we present a construction of Noga Alon which shows that our analysis is tight amongst a certain class of approaches.

2.1 Spread boosting

The following is the analog of the spread-boosting theorem in [GLR08], except we only care about the mass outside edges in *induced* subgraphs of G (and not an arbitrary collection of edges of certain size).

Lemma 2. *Let $G = (V, E)$ be an (n, d, λ)-expander, and let $L \subseteq \mathbb{R}^d$ be a $(d/B, \rho)$-spread subspace for some parameters $B > 1$ and $\rho < 1$. Then, for all $W \subseteq V$, $|W| \leqslant \frac{n}{2B}$, there exists a subset $Z \subseteq W$, $|Z| \leqslant \left(\frac{2\lambda B}{d}\right)^2 |W|$ such that for every $x \in T(G, L)$ the following holds:*

$$\sum_{e \notin E(W)} x_e^2 \geqslant \rho^2 \sum_{e \notin E(Z)} x_e^2 \ . \tag{2}$$

Proof. Given W, we define Z as follows:

$$Z = \left\{ w \in W \ : \ |N_G(w) \cap W| \geqslant \frac{d}{B} \right\} \ .$$

By definition, $|E(Z,W)| \geqslant \frac{d}{B}|Z|$. On the other hand, by the expander mixing lemma (see, e.g. [HLW06, §2.4]),

$$|E(Z,W)| \leqslant d|Z|\frac{|W|}{n} + \lambda\sqrt{|Z||W|} \leqslant \frac{d|Z|}{2B} + \lambda\sqrt{|Z||W|} \ .$$

Combining the two bounds, $|Z| \leqslant \left(\frac{2\lambda B}{d}\right)^2 |W|$. By definition of Z and the $(d/B, \rho)$-spread property of L, it follows easily that

$$\sum_{e \in E(W,\overline{W})} x_e^2 \geqslant \rho^2 \sum_{v \in W \setminus Z} \|x_{\Gamma(v)}\|_2^2 \ . \tag{3}$$

Now

$$\sum_{v \in W \setminus Z} \|x_{\Gamma(v)}\|_2^2 \geqslant \|x\|_2^2 - \sum_{e \in E(\overline{W})} x_e^2 - \sum_{e \in E(Z)} x_e^2 = \sum_{e \notin E(Z)} x_e^2 - \sum_{e \in E(\overline{W})} x_e^2 \ . \tag{4}$$

Combining the previous two bounds, we get the desired conclusion (2).

Remark 1 (Comparison to [GLR08]). A generalization of the $T(G,L)$ construction (and, indeed, the natural setting for expander codes) is to consider a *bipartite* graph $H = (V_L, V_R, E)$ with $N = |V_L|$, in which every node of V_R has degree exactly d. In this case, given $L \subseteq \mathbb{R}^d$, one defines the subspace

$$X(H,L) = \left\{x \in \mathbb{R}^N : x_{\Gamma_H(j)} \in L \text{ for every } j \in V_R\right\},$$

where now $\Gamma_H(j) \subseteq V_L$ denotes the neighbors of a vertex $j \in V_R$. Clearly $T(G,L) = X(H,L)$, where H is defined by $V_L = E(G)$, $V_R = V(G)$, and the edges of H are naturally given by edge-vertex incidence in G.

The paper [GLR08] analyzes the well-spreadness of $X(H,L)$ in terms of the well-spreadness of L and the combinatorial expansion properties of H (i.e. how large are the neighbor sets of subsets $S \subseteq V_L$). There does not exist a bipartite expander graph H with properties strong enough to imply our main result (Corollary 7), if one requires expansion from *every* subset $S \subseteq V_L$, and uses only the iterative spreading analysis of [GLR08]. We overcome this by structuring the iteration so that only certain subsets arise in the analysis, and we only require strong expansion properties from these subsets. Lemma 2 represents this idea, applied to the subspace $T(G,L)$. Here, the special subsets are precisely those edge sets which arise from induced subgraphs of G (as opposed to arbitrary subsets of the edges).

Iterating Lemma 2 yields the following.

Corollary 5 *Let $G = (V,E)$ be an (n,d,λ)-expander, and $L \subseteq \mathbb{R}^d$ be a $(d/B, \rho)$-spread subspace. Let ℓ be an integer such that*

$$\left(\frac{d}{2\lambda B}\right)^{2\ell} \geqslant \frac{n}{2B} \ .$$

Then for all $x \in T(G,L)$ with $\|x\|_2 = 1$ and every $W \subseteq V$ with $|W| < \frac{n}{2B}$, we have

$$\sum_{e \notin E(W)} x_e^2 \geqslant \rho^{2\ell} \ .$$

We now come to the proof of our main theorem.

Theorem 6 (Main). *Let $G = (V,E)$ be an (n,d,λ)-expander with $N = nd/2$ edges. Let $L \subseteq \mathbb{R}^d$ be a $(d/B,\rho)$-spread subspace of co-dimension at most ηd for some parameters $\rho, \eta < 1$ and $B > 1$. Then the subspace $T(G,L) \subseteq \mathbb{R}^N$ has dimension at least $N(1-2\eta)$ and it is*

$$\left(\frac{N}{2B^2}, \frac{\rho}{\sqrt{2}} n^{\frac{-\log(1/\rho)}{\log(d/(2\lambda B))}}\right) \text{-spread} \ .$$

In particular, by Lemma 1, this implies

$$\Delta(T(G,L)) \leqslant \frac{2\sqrt{2}B}{\rho^2} \cdot n^{\frac{2\log(1/\rho)}{\log(d/(2\lambda B))}} \ .$$

Proof. The claim about the dimension of $T(G,L)$ is obvious. Fix an arbitrary $x \in T(G,L)$ with $\|x\|_2 = 1$. Let $F \subseteq E$ be an arbitrary set with $|F| \leqslant \frac{N}{2B^2} = \frac{nd}{4B^2}$. We need to prove that

$$\sum_{e \notin F} x_e^2 \geqslant \frac{\rho^2}{2} n^{\frac{-2\log(1/\rho)}{\log(d/(2\lambda B))}} \ . \tag{5}$$

Define

$$W = \left\{ v \in V \ : \ |F \cap \Gamma(v)| > d/B \right\} . \tag{6}$$

Since $2|F| > |W|d/B$, we have $|W| < \frac{2|F|B}{d} \leqslant \frac{n}{2B}$. We can now apply Corollary 5 with a choice of ℓ that satisfies $\left(\frac{d}{2\lambda B}\right)^{2\ell} \leqslant nd \leqslant n^2$, and conclude that

$$\sum_{e \notin E(W)} x_e^2 \geqslant \rho^{2\ell} \geqslant n^{-\frac{2\log(1/\rho)}{\log(d/(2\lambda B))}} \tag{7}$$

We have the chain of inequalities

$$2\sum_{e \notin F} x_e^2 = \sum_{v \in V} \|x_{\Gamma(v)\setminus F}\|_2^2 \geqslant \sum_{v \notin W} \|x_{\Gamma(v)\setminus F}\|_2^2 \geqslant \rho^2 \sum_{v \notin W} \|x_{\Gamma(v)}\|_2^2 \geqslant \rho^2 \sum_{e \notin E(W)} x_e^2 \ , \tag{8}$$

where the last but one step follows since L is $(d/B,\rho)$-spread, and $|\Gamma(v) \cap F| \leqslant d/B$ when $v \notin W$. Combining (7) and (8) gives our desired goal (5).

The main application of the above theorem is the following result.

Corollary 7 (Constant distortion with sub-linear randomness) *For every $\delta, \eta > 0$ and every $N \geq 1$, there is a randomized (Monte Carlo) construction of a subspace $X \subseteq \mathbb{R}^N$ using N^δ random bits that has dimension at least $(1-\eta)N$ and distortion $\Delta(X) \leqslant \left(\frac{1}{\eta}\right)^{O(1/\delta)}$.*

Proof. For every $N' \geq 1$, it is known how to construct an explicit (n, d, λ)-expander, with $\lambda \leq d^{0.9}$, $n^{\delta/4} \leq d \leq n^{\delta/2}$ such that $N' \leq N \leq 10N'$, where $N = nd/2$ [LPS88] (also see [HLW06, §2.6.3] for a discussion of explicit constructions of expander graphs).

Let $L \subset \mathbb{R}^d$ be the kernel of a random $(\eta/2)d \times d$ sign matrix. Constructing L requires at most n^δ random bits, and by Theorem 2, L is $(\eta^{O(1)}d, \eta^{-O(1)})$-spread with high probability. When this happens, the subspace $T(G, L) \subseteq \mathbb{R}^N$ has distortion at most $(\frac{1}{\eta})^{O(1/\delta)}$ by Theorem 6.

The above description only works for (infinitely many) values N of a certain form. With some combinatorial manipulations, it can be made to work for all N; see [GLR08, §2.2.2].

2.2 Optimality of myopic analysis

The analysis of the previous section is myopic in the sense that it only cares about the expansion properties of G, and the spreading properties of the local subspace L. The following construction, suggested to us by Noga Alon, shows that if we only use the fact that G is an expander, and that every vector induced on the neighbors of a vertex is well-spread, then asymptotically our analysis is tight. The point is that, unlike in the boolean setting, real numbers come in scales, allowing them to decay slowly while still satisfying the local constraints.

Theorem 8 ([Alo08]). *For every $d \geq 4$, and infinitely many $n \in \mathbb{N}$, there exists an $(n, d, O(\sqrt{d}))$-expander $G = (V, E)$ with $N = |E|$, and a point $x \in \mathbb{R}^N$, $x \neq 0$, such that*

$$\|x_{\Gamma(v)}\|_1 \geq \frac{\sqrt{d}}{2}\|x_{\Gamma(v)}\|_2$$

for every $v \in V$, but $\|x\|_1 \leq N^{\frac{1}{2}-\Omega(\frac{1}{\log d})}\|x\|_2$.

Proof. Let $n = 2(d-1)^k$ for some $k \in \mathbb{N}$, and let H be an $(n, d-1, O(\sqrt{d}))$-expander. Let T' and T'' be two disjoint, complete $(d-1)$-ary rooted trees of depth k. Let T be the tree that results from adding an edge e_0 between the roots of T' and T''. Finally, define $G = (V, E)$ as the d-regular graph that results from identifying the n leaves of T in an arbitrary way with the nodes of H. It is easy to check that G is an $(n', d, O(\sqrt{d}))$-expander with $n' = \Theta(d^k)$ and $N = |E| = \Theta(d^{k+1})$.

We may think of $x \in \mathbb{R}^N$ as indexed by edges $e \in E$. For $e \in E$, let $h(e)$ be the distance from e to e_0, and put $x_e = (2\sqrt{d})^{-h(e)}$. It is straightforward to verify that, for every $v \in V$, one has

$$\frac{\|x_{\Gamma(v)}\|_1}{\|x_{\Gamma(v)}\|_2} = \frac{2\sqrt{d} + d - 1}{\sqrt{5d-1}} \geq \frac{\sqrt{d}}{2}.$$

Clearly we have $\|x\|_2 \geq 1$, whereas

$$\|x\|_1 = 1 + 2\left[\sum_{h=1}^{k}\left(\frac{d-1}{2\sqrt{d}}\right)^h\right] + \left(\frac{d-1}{2\sqrt{d}}\right)^{k+1}$$

$$= O(2^{-k} d^{(k+1)/2}) = O(2^{-k}\sqrt{N}) \leq N^{\frac{1}{2} - \Omega(\frac{1}{\log d})}.$$

We remark that the vector $x \in \mathbb{R}^N$ exhibited in the preceding theorem lies in $T(G, L)$, where $L = \text{span}(1, \frac{1}{2\sqrt{d}}, \ldots, \frac{1}{2\sqrt{d}})$, as long as we choose the ordering of the neighborhoods $\Gamma(v)$ appropriately (recall that one has to fix such an ordering in defining $T(G, L)$). In light of this obstruction to our analysis, the following question seems fundamental.

Question: Is there a number $K \in \mathbb{N}$ such that for infinitely many N, there exists an $\frac{N}{2} \times N$ $\{0,1\}$-matrix A, with at most K ones per row, and such that $\Delta(\ker(A)) = O(1)$?

Theorem 6 shows that one can take $K = N^{\delta}$ for any $\delta > 0$, but this is the best current bound.

3 Error-correction over reals

In this section, we discuss connections of our work to compressed sensing, or equivalently to error-correcting codes over reals. We will use the coding terminology to describe these connections.

3.1 Background

An w-error-correcting code of dimension m and block length N over the reals is given by a linear map $C : \mathbb{R}^m \to \mathbb{R}^N$, such that for each $f \in \mathbb{R}^m$, $f \neq 0$, $\|Cf\|_0 > 2w$. The rate of the code is the ratio m/N. Given a received word $y = Cf + e$ with $\|e\|_0 \leqslant w$, one can recover the message f as the solution x to optimization problem:

$$\min_{x \in \mathbb{R}^m} \|y - Cx\|_0 \ .$$

The above non-convex optimization problem is NP-hard to solve in general. Quite remarkably, if the code C meets certain conditions, one can recover f by solving the linear program (LP)

$$\min_{x \in \mathbb{R}^m} \|y - Cx\|_1 \ .$$

(The above ℓ_1-minimization task, which is easily written as a linear program, is often called *basis pursuit* in the literature.) Note that we are not restricting the magnitude of erroneous entries in e, only that their number is at most w.

Candes and Tao [CT05] studied the above error-correction problem and proved that ℓ_1-minimization works if the code has a certain restricted isometry property. A sufficient condition for ℓ_1-minimization to work is also implied by the distortion property of the image of C, as formalized in the following lemma by Kashin and Temlyakov [KT07].

Lemma 3 ([KT07]). *Let $X = \{Cx \mid x \in \mathbb{R}^m\} \subseteq \mathbb{R}^N$ be the image of C. Then C is a w-error-correcting code provided $w < \frac{N}{4\Delta(X)^2}$, and moreover, given $y \in \mathbb{R}^N$ such that $\|y - Cf\|_0 \leqslant w$ for some $f \in \mathbb{R}^m$, the signal f can be recovered efficiently by solving the (LP)* $\min_{x \in \mathbb{R}^m} \|y - Cx\|_1$.

By picking X to be the kernel of a random $\gamma N \times N$ sign matrix, and plugging in the distortion bound of Theorem 2, gives codes of rate at least $(1-\gamma)$ that are, w.h.p, w-error-correcting with an efficient algorithm for $w = \Theta(\frac{\gamma N}{\log(1/\gamma)})$. This is not far from the best possible bound, achieved non-constructively, of $w = \Theta(\gamma N)$.

The ℓ_1-minimization decoding algorithm, while polynomial time, requires solving a dense linear program of size N. It is of interest to develop faster methods for decoding, ideally with near-linear complexity.

3.2 Near-linear time decoding

Before we begin, let us remark that in the compressed sensing setting, since $T(G, L)$ is the kernel of a relatively sparse matrix A (with at most N^δ non-zero entries per row), the sensing, i.e., the computation of Ax for $x \in \mathbb{R}^N$, can be done in $O(N^{1+\delta})$ time. This sparsity also makes interior point methods for basis pursuit more efficient by a similar factor. Note that in the language of codes, "sensing" corresponds to syndrome computation, while signal recovery (recovering x from a noisy version of Ax) corresponds to syndrome decoding.

We now turn to algorithms for our construction $X = T(G, L)$. Efficient encoding follows immediately from sparsity of the encoding matrix (in fact, we achieve linear-time encoding by choosing $d = O(1)$ in what follows). Near-linear time decoding will be achieved using a natural iterative algorithm for Tanner codes.

For technical reasons that help with the decoding, we will take $G = (V_\ell, V_r, E)$ to be a d-regular $n \times n$ *bipartite* graph. Specifically, we will take G to be the double cover of an (n, d, λ)-expander, and $L \subseteq \mathbb{R}^d$ to be the kernel of a random $\gamma d/2 \times d$ matrix (for a constant $\gamma > 0$). With this choice, $X \subseteq \mathbb{R}^N$ satisfies $\mathsf{dim}(X) \geqslant (1-\gamma)N$ where $N = nd$, and our code has rate at least $(1-\gamma)$. With high probability, L will be ζd-error-correcting via ℓ_1-minimization for $\zeta = \Theta(\frac{\gamma}{\log(1/\gamma)})$, and we will assume this is the case.

There is a natural iterative algorithm for Tanner codes [SS96], specifically for the version when the underlying graph is bipartite [Zól] (see also [GI05, Sec. 2.2]), which can be adapted in a straightforward manner to the coding over reals setting. An adaptation of the related sequential "bit-flipping" algorithm for expander based low-density parity-check codes appears in [XH07]; using lossless expanders, their approach also leads to a near-linear time decoding algorithm.

Our algorithm for decoding $T(G, L)$ proceeds in several rounds, alternating between "left" and "right" rounds. In a left round, on input a string $y \in \mathbb{R}^N$ from the previous decoding round (or the noisy codeword at the start of decoding),

we locally decode $y_{\Gamma(u)}$ for each $u \in V_\ell$ to its closest vector in L in ℓ_1-sense. In the subsequent "right" round, we do the same for each $v \in V_r$, and then switch back to a left round. The key point is that if the number of errors in the local neighborhood $y_{\Gamma(u)}$ of a vertex is less than ζd, then the local ℓ_1-minimization will correct those errors and thus fix this local neighborhood. The decoder terminates when either all the local projections $y_{\Gamma(u)}$, $u \in V_\ell \cup V_r$, belong to L (so that globally we have decoded to a codeword of $T(G, L)$), or more than $\Omega(\log N)$ iterations have passed without convergence to a global codeword (in this case, there must have been too many errors in the original input).

Only $O(\log N)$ iterations suffice because the number of local neighborhoods which are not yet fully decoded decays geometrically in each round, as long as the initial number of errors is small enough. Arguing as in [Z01], one can show the following.

Theorem 9. *If G is the double cover of an (n, d, λ)-expander with $\lambda \leq \gamma^2 d$, and $N = nd$, then the above algorithm decodes $T(G, L)$ and corrects up to $w = O\Big(\frac{\gamma^2}{\log^2(1/\gamma)} N\Big)$ errors. Further, the algorithm runs in $O(Nt(d) \log_d N)$ time (or in $O(t(d) \log_d N)$ parallel time with $O(N)$ processors), where $t(d)$ is the time to perform ℓ_1-minimization for the subspace $L \subseteq \mathbb{R}^d$ (and is thus a constant if d is a constant).*

Thus if we settle for a slightly worse fraction of errors, namely $w/N = \Theta\Big(\frac{\gamma^2}{\log^2(1/\gamma)}\Big)$ instead of $\Theta\Big(\frac{\gamma}{\log(1/\gamma)}\Big)$, then the decoding can be performed in near-linear time.

An argument similar to the one in Section 2.1 can be used to show that $\Delta(X) \leqslant N^{O(\frac{\log(1/\gamma)}{\log d})}$. In fact, the sequence of sets that arise in the repeated application of Lemma 2, starting from the set W defined in (6), would correspond to the subsets of vertices, alternating between the left and right sides, that arise in decoding a vector supported on $F \subseteq E$ to the all-zeroes vector. Note that by the connection mentioned in Lemma 3, this would only enable correcting w errors for $w \leqslant N^{1-\Omega(1/\log d)}$ via global ℓ_1-minimization, compared to the $\Omega(N)$ errors handled by the iterative decoder.

However, there is a substantial shortcoming of the error model used in Theorem 9. In practice, it is not reasonable to assume that the error is only supported on w positions. It is more reasonable to allow small non-zero noise even in the remaining positions (recall that we assume no bound on the magnitude of noise in the w erroneous positions); this error model is used in previous works like [CT05]. The ℓ_1-minimization works also in this setting; specifically, if $w \leqslant O(N/\Delta(X)^2)$ and the error vector e satisfies $\|e - \sigma_w(e)\|_1 \leqslant \varepsilon$ where $\sigma_w(e)$ is the vector with the w largest components of e and rest equal to zero, then ℓ_1-minimization recovers a string z such that $\|z - Cf\|_2 \leqslant \frac{\Delta(X)}{\sqrt{N}}\varepsilon$ (see [KT07]).

Extending iterative decoding to the above setting is an interesting challenge that we hope to study in future work.

Acknowledgments

We are grateful to Noga Alon for his help with the proof of Theorem 8 and his permission to include it here.

References

[AAM06] 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.

[Alo08] Noga Alon. Personal communication, 2008.

[CT05] Emmanuel J. Candes and Terence Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[Don06] David L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006.

[FLM77] 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.

[GG84] A. Garnaev and E. D. Gluskin. The widths of Euclidean balls. *Doklady An. SSSR.*, 277:1048–1052, 1984.

[GI05] Venkatesan Guruswami and Piotr Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Trans. Inform. Theory*, 51(10):3393–3400, 2005.

[GLR08] V. Guruswami, J. R. Lee, and A. Razborov. Almost Euclidean subspaces of ℓ_1^N via expander codes. In *SODA '08: Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 353–362, Philadelphia, PA, USA, 2008. Society for Industrial and Applied Mathematics.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.

[Ind06] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM*, 53(3):307–323, 2006.

[Ind07] Piotr Indyk. Uncertainty principles, extractors, and explicit embeddings of L_2 into L_1. In *Proceedings of the 39th Annual ACM Symposium on the Theory of Computing*, pages 615–620, 2007.

[JS01] William B. Johnson and Gideon Schechtman. Finite dimensional subspaces of L_p. In *Handbook of the geometry of Banach spaces, Vol. I*, pages 837–870. North-Holland, Amsterdam, 2001.

[Kas77] 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.

[KT07] B. S. Kashin and V. N. Temlyakov. A remark on compressed sensing. Available at `http://www.dsp.ece.rice.edu/cs/KT2007.pdf`, 2007.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[LS07] 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.

[Mil00] V. Milman. Topics in asymptotic geometric analysis. *Geom. Funct. Anal.*, (Special Volume, Part II):792–815, 2000. GAFA 2000 (Tel Aviv, 1999).

[Sch84] Carsten Schütt. Entropy numbers of diagonal operators between symmetric Banach spaces. *J. Approx. Theory*, 40(2):121–128, 1984.

[SS96] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inform. Theory*, 42(6, part 1):1710–1722, 1996. Codes and complexity.

[Sza06] Stanislaw Szarek. Convexity, complexity, and high dimensions. In *International Congress of Mathematicians. Vol. II*, pages 1599–1621. Eur. Math. Soc., Zürich, 2006.

[Tan81] Robert M. Tanner. A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5):533–547, 1981.

[XH07] Weiyu Xu and Babak Hassibi. Efficient compressive sensing with determinstic guarantees using expander graphs. In *IEEE Information Theory Workshop*, September 2007.

[Zó1] Gillés Zémor. On expander codes. *IEEE Transactions on Information Theory*, 47(2):835–837, 2001.