Expander codes over reals, Euclidean sections, and compressed sensing

Venkatesan Guruswami, James R. Lee, and Avi Wigderson

***Abstract*— Classical results from the 1970's state that w.h.p. a random subspace of N-dimensional Euclidean space of proportional (linear in N) dimension is "well-spread" in the sense that vectors in the subspace have their ℓ_2 mass smoothly spread over a linear number of coordinates. Such well-spread subspaces are intimately connected to low distortion embeddings, compressed sensing matrices, and error-correction over reals.**

We describe a construction inspired by expander/Tanner codes that can be used to produce well-spread subspaces of $\Omega(N)$ dimension using sub-linear randomness (or in sub-exponential time). These results were presented in our paper [10]. We also discuss the connection of our subspaces to compressed sensing, and describe a near-linear time iterative recovery algorithm for compressible signals.

I. Introduction

The topic broadly referred to as "compressed sensing" has been intensively studied in recent years.[1] In one formulation of compressed sensing, the goal is to approximately recover a high-dimensional vector $x \in \mathbb{R}^N$ based on its low-dimensional sketch $Ax \in \mathbb{R}^m$ (for $m \ll N$), where A is an $m \times N$ "sensing" matrix. (The vector Ax is a *linear* compression of the real signal x.) While such recovery is in general impossible to achieve (since $Ax = y$ is an undetermined linear system with many solutions), it becomes possible when the signal x is sufficiently sparse (or close to sparse). Specifically, it is known that when A is drawn at random from certain ensembles (eg. with i.i.d Gaussians or random ± 1 entries), given $z = Ax$, the solution $x^\#$ to the ℓ_1-minimization problem:

$$\text{Minimize } \|x'\|_1 \text{ subject to } Ax' = z \ , \tag{1}$$

is close to the best k-sparse approximation for the signal x for $k \lesssim m/\log(N/m)$. Formally, one has the error guarantee

$$\|x - x^\#\|_2 \leqslant O\Big(\frac{1}{\sqrt{k}}\Big) \min_{y \in \mathbb{R}^N, \|y\|_0 \leqslant k} \|x - y\|_1 \ . \tag{2}$$

(Here $\|y\|_0$ refers to the number of nonzero coordinates in y.) Note that the above is equivalent to

$$\|x - x^\#\|_2 \leqslant O\Big(\frac{1}{\sqrt{k}}\Big) \|x - \sigma_k(x)\|_1 \ ,$$

where $\sigma_k(x)$ is the vector with the k largest entries of x in absolute value, and the rest of the entries equal to 0.

(1) is a convex optimization problem (it is in fact a linear program) and can be solved in polynomial time, so we also get a polynomial time signal recovery algorithm. The approximation guarantee (2) is a "mixed" ℓ_2-ℓ_1 guarantee. The ℓ_1 minimization algorithm also achieves an ℓ_1-ℓ_1 error guarantee

$$\|x - x^\#\|_1 \leqslant O(1) \cdot \|x - \sigma_k(x)\|_1 \ . \tag{3}$$

The ℓ_1-ℓ_1 guarantee is in general weaker than the ℓ_2-ℓ_1 approximation (2). We can only show the ℓ_1-ℓ_1 guarantee for the signal recovery algorithm we give in this paper.

A natural goal in the context of compressed sensing is to find properties on the sensing matrix A that suffice for approximately recovering sparse signals, say with guarantees of the form (2), in polynomial time. Candes, Romberg, and Tao [6], [5] established that a criterion for matrices called "Restricted Isometry Property" (RIP) implies that ℓ_1-minimization on the sketch Ax outputs a signal $x^\#$ satisfying (2). The RIP property in turn is closely related to, and in fact implied by, the classic Johnson-Lindenstrauss dimensionality reduction property [3].

The RIP property, however, seems stronger than what is needed for the compressed sensing property (2). A different route to obtain good compressed sensing matrices is via classical results in high-dimensional geometry on *near Euclidean sections* of the ℓ_1 ball [7], [12], [8]. These results state that if A is a random $m \times N$ matrix, say with i.i.d Gaussian entries, then the subspace $\ker(A)$ has *distortion* $\lesssim \sqrt{\frac{N}{m}\log\frac{N}{m}}$, where the distortion of a subspace $X \subseteq \mathbb{R}^N$ is defined to be

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

Note that for every $X \subseteq \mathbb{R}^N$, $1 \leqslant \Delta(X) \leqslant \sqrt{N}$. A small value of $\Delta(X)$ means that every $x \in X$ is "well-spread" in the sense that its ℓ_2-mass cannot be concentrated on a small number of coordinates (see, eg. [9], for a precise quantitative connection between distortion and ℓ_2-spread properties).

In a beautiful paper, Kashin and Temlyakov [13] make this connection precise — they prove that if $\ker(A)$ has distortion Δ, then A is a compressed sensing matrix for which ℓ_1-minimization (1) provides the approximation guarantee (2) for sparsity parameter $k \approx N/\Delta^2$. Plugging in the optimal distortion bound of $\Delta \lesssim \sqrt{\frac{N}{m}\log\frac{N}{m}}$ one gets that k can be as large as $\Omega(m/\log(N/m))$ (matching the RIP based bounds). The authors of [13] also show a weak converse stating that approximate sparse signal recovery with the

Venkatesan Guruswami is with the Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213. Email: guruswami@cmu.edu. Research supported by NSF CCF-0835814 and a Packard Fellowship.

James Lee is with the Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195. Email: jrl@cs.washington.edu. Research supported by NSF CCF-0644037.

Avi Wigderson is with the School of Mathematics, Institute for Advanced Study, Princeton, NJ 08540. Email: avi@ias.edu. Research supported by NSF grants CCF-0832797 and DMS-0835373.

[1]See http://dsp.rice.edu/cs for a comprehensive collection of references on compressive sensing.

strong ℓ_2-ℓ_1 guarantee (which they call strong compressed sensing property) implies that $\ker(A)$ has low distortion.

Based on the above discussion, there are various angles to reason about matrices with good compressed sensing guarantees. Yet, all these results on RIP matrices or low distortion subspaces are probabilistic, and show that the desired property holds with high probability when a matrix is picked randomly from a suitable ensemble. Once such a matrix is sampled, it is not known how to efficiently verify if it has the desired restricted isometry, distortion, or compressed sensing properties.

A natural and important question, therefore, is whether one can *explicitly* (in deterministic polynomial time) construct a matrix with good compressed sensing properties. An explicit construction is desirable as it would provide a certified guarantee of the desired property, lead to a better understanding of the underlying structure, and also potentially result in more efficient algorithms (that exploit some feature of the construction, such as sparsity). Such explicit constructions are, however, quite challenging, and seem much harder to come by.

In this work, we will address the explicit construction question for the problem of constructing low distortion subspaces of $\mathbb{R}^N$ of proportional dimension $\Omega(N)$. When the dimension d of the subspace is $\Theta(N)$, the results of [12] state that a random subspace of dimension d, defined by the kernel of a $(N-d)\times N$ matrix with i.i.d Gaussian (or random ± 1) entries, has distortion $O(1)$. This gives a construction of such subspaces using $O(N^2)$ random bits. This randomness requirement was reduced to $O(N\log N)$ in [2] and later to $O(N)$ in [14]; due to an underlying union bound over a large ε-net, these approaches are unlikely to yield constructions with sub-linear randomness. We give a derandomized construction of such subspaces with distortion $O_\delta(1)$ using only N^δ random bits, for any desired $\delta > 0$. This implies a deterministic construction in $2^{O(N^\delta)}$ time. Unfortunately, our distortion bound, though a constant for every fixed $\delta > 0$, is exponential in $1/\delta$.

We will also analyze a natural iterative signal recovery algorithm when matrix we construct is used for compressed sensing. The advantage of the algorithm over ℓ_1-minimization (1) is that it runs in near-linear time. Further, we are able to prove a better bound on the sparsity up to which the algorithm works (compared to via the connection to distortion [13] mentioned previously). However, the algorithm only provides the ℓ_1-ℓ_1 approximation guarantee (3), and we do not know how to achieve an approximation bound like (2). We remark that other recent combinatorial signal recovery algorithms in the context of compressed sensing such as expander/sparse matching pursuit [11], [4] also only provide an ℓ_1-ℓ_1 approximation guarantee. We now turn to the formal statements of our results.

II. EUCLIDEAN SECTIONS OF ℓ_1^N FROM EXPANDER CODES

Tanner [16] gave an approach to construct long codes C from a short code C', by imposing a collection of "local" conditions that projections of codewords in C on certain subsets of coordinates belong to C'. The structure of the local conditions can be captured by a bipartite graph which is referred to as the factor graph. Below we use the continuous version of this approach to construct a large global subspace X of $\mathbb{R}^N$ by combining a "local" subspace L of $\mathbb{R}^d$ for $d \ll N$ with a d-regular graph. (This construction originally appears in our paper [10].) Each vector x in X must satisfy the local conditions that the projection of x on the neighborhood of each vertex belongs to L. When instantiated with the appropriate local subspace and graph, this allows us to construct large subspaces of good distortion from a near-optimal small (local) subspace. In the context of compressed sensing, this allows us to improve the sparsity up to which signals can be recovered as well as the efficiency of signal recovery.

Definition 1: 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

$$T(G,L) = \{x \in \mathbb{R}^N \mid x_{E(v)} \in L \text{ for every } v \in V\} \ . \quad (4)$$

where $E(v)$ is the set of d edges incident on v in some fixed order.

This is the Tanner construction when the factor graph is the edge-vertex incidence graph of the d-regular graph G. A similar construction of codes was given and analyzed by Sipser and Spielman [15], and the above is the continuous analog of this construction. The expander codes based approach to constructing subspaces was initiated in [9] where it was used, along with several other ingredients, to construct an explicit subspace of dimension $N - o(N)$ and distortion $\approx (\log N)^{\log\log\log N}$.

We will use the spectral definition of graph expansion, as defined below.

Definition 2 (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.

The following theorem is proved in [10].

Theorem 1: Let G be an (n,d,λ)-expander for $\lambda \leqslant d^{0.9}$, and $L \subseteq \mathbb{R}^d$ be a subspace of dimension $\dim(L) \geqslant (1-\eta)d$ and distortion $\Delta(L) \leqslant c_0$ for some absolute constants $c_0 > 1$ and $0 < \eta < 1/2$. Then $T(G,L) \subseteq \mathbb{R}^N$ (where $N = nd/2$) satisfies $\dim(T(G,L)) \geqslant (1-2\eta)N$ and $\Delta(T(G,L)) \leqslant N^{O(1/\log d)}$ (the constant in $O(1/\log d)$ depends on c_0).

By picking $d = n^{\delta/2}$, choosing L to be the kernel of an $\eta d \times d$ matrix with random and independent ± 1 entries, and taking G to be a Ramanujan expander of degree d (and second largest eigenvalue $\lambda = O(\sqrt{d})$), one can conclude the following:

Corollary 2: For every $\delta > 0$, for infinitely many integers N, there is a construction of a subspace X of $\mathbb{R}^N$ with $\dim(X) \geqslant N/2$ and $\Delta(X) \leqslant 2^{O(1/\delta)}$ that uses N^δ random bits. Such a subspace X can also be constructed in deterministic $2^{O(N^\delta)}$ time.

The deterministic construction part follows since one can construct the required local subspace in $\mathbb{R}^d$ deterministically in $2^{O(d^2)}$ time by going over all sign matrices and estimating their distortion using a sufficiently dense net for the unit ball.

III. Using our subspaces for compressed sensing

We now turn to the compressed sensing properties of the above Tanner construction of subspaces, or more precisely of the matrix, say A, whose kernel is $T(G, L)$. We find it slightly more convenient to present these results in the terminology of error-correction over reals, where the goal is to decode an input string in $\mathbb{R}^N$ to a close-by vector in $T(G, L)$ (when a close enough such vector exists). Formally, given $y = c + e$, where $c \in T(G, L)$ and e is a sparse (or nearly sparse) error vector, the goal is to recover an approximation to c. Compressed sensing corresponds to "syndrome decoding" where one is given the syndrome $Ay = Ae$ and the goal is to recover (an approximation) to e. The sparsity parameter in compressed sensing corresponds to the number of errors (of arbitrary magnitude) that can be corrected.

Given the distortion bound of Theorem 1, the connection of Euclidean sections to compressed sensing proved by Kashin and Temlyakov [13] implies that given $y = c + e \in R^N$ where $c \in T(G, L)$, using ℓ_1-minimization one can find $c' \in T(G, L)$ such that $\|c' - c\|_2 \leqslant O(\frac{1}{\sqrt{t}})\|e - \sigma_t(e)\|_1$ for $t = N^{1-O(1/\log d)}$. Thus one can recover a good approximation of the "codeword" c provided the error vector has small ℓ_1-mass outside its $\approx N^{1-O(1/\log d)}$ largest coordinates.

This bound is rather weak, and for the parameters of Corollary 2, the fraction of errors will be $\exp(-1/\delta)$. We now proceed to improve this bound on the number of errors t of arbitrary magnitude that can be corrected. This will be based on an iterative algorithm, based on a sequence of local decoding steps. This is the continuous analog of the Sipser-Spielman-Zémor approach for decoding expander based binary Tanner codes [15], [18]. Xu and Hassibi [17] present an iterative algorithm for recovering *exactly* sparse signals for a subspace construction based on unbalanced bipartite lossless expanders (this construction is the continuous analog of Gallager's LDPC codes).

The main subtlety in our algorithm (which works for all signals, not just exactly t-sparse signals) is to control the accumulation of error outside the t largest coordinates since we are no longer in a discrete setting where we can assume that all but the largest t coordinates equal 0. Our formal result is the following. For some technical reasons, we will use the construction $T(G, L)$ for a *bipartite* d-regular graph G. Such a graph can be constructed by taking the "double cover" of a d-regular undirected graph H.[2]

Theorem 3 (Main): Consider the subspace $T(G, L) \subset \mathbb{R}^{nd}$ where G is the double cover of $(n, d, \lambda = O(\sqrt{d}))$-expander and $L \subseteq \mathbb{R}^d$ is a subspace with $\dim(L) \geqslant (1-\eta)d$ and $\Delta(L) \leqslant c_0$ for some absolute constants $\eta < 1/2$ and $c_0 > 1$. Let $N = nd$. There is a constant $\beta = \beta(c_0) > 0$ and a deterministic signal recovery algorithm $\mathcal{A}$ running in time $O(Nd^{O(1)} \log_d N)$ that provides the following guarantee: Given as input a string $\mathbf{y} \in \mathbb{R}^N$ expressible as $\mathbf{y} = \mathbf{c} + \mathbf{f}$ for some $\mathbf{c} \in T(G, L)$ and $\mathbf{f} \in \mathbb{R}^N$, the algorithm $\mathcal{A}$ outputs a $\mathbf{c}' \in \mathbb{R}^N$ (but not necessarily in $T(G, L)$) satisfying

$$\|\mathbf{c}' - \mathbf{c}\|_1 \leqslant 2^{O(\log_d n)}\|\mathbf{f} - \sigma_{\beta N}(\mathbf{f})\|_1 \ . \tag{5}$$

Picking $d = n^\delta$ for a constant $\delta > 0$ in the above leads to an algorithm running in $N^{1+O(\delta)} \cdot O(1/\delta)$ time with an ℓ_1-ℓ_1 approximation guarantee (3) with a multiplicative constant $2^{O(1/\delta)}$. The fraction of errors corrected, or equivalently the sparsity parameter for compressed sensing, is a constant fraction β of the signal length N. We note that β depends only on the distortion c_0 of the local subspace L, and is *independent* of δ (unlike the ℓ_1-minimization algorithm where it was of the form $2^{-\Omega(1/\delta)}$).

In the remainder of this section, we will describe the algorithm $\mathcal{A}$ and prove Theorem 3.

Let $L = \{x \in \mathbb{R}^d \mid Mx = 0\}$ be a subspace of $\mathbb{R}^d$ defined as the kernel of a matrix $M \in \mathbb{R}^{\eta d \times d}$. By the previous discussion on the connection between distortion and compressed sensing [13], we can assume that L has the following error-correction guarantee for

$$w = \alpha d \quad \text{where} \quad \alpha = \Theta(1/c_0^2) \ . \tag{6}$$

Given $z = x + e$ for $x \in L$, the solution z to the linear program:

Minimize $\|z'\|_1$ subject to $Mz' = Mz$

satisfies

$$\|z' - e\|_1 \leqslant C \cdot \|e - \sigma_w(e)\|_1 \tag{7}$$

for some absolute constant $C > 1$. In other words, the "codeword" $z - z' \in L$ is close to the correct codeword x in ℓ_1-sense, i.e., $\|(z-z')-x\|_1 \leqslant C \cdot \varepsilon$, provided the error vector e has ℓ_1-mass at most ε outside its w most significant coordinates.

We will now prove Theorem 3 for the choice

$$\beta = \alpha^2/2 \ .$$

Set $t = \beta N$ to denote the number of errors of arbitrary magnitude that we wish to correct. We will use boldface $\mathbf{y}$ to denote vectors in $\mathbb{R}^N$ to distinguish them from local vectors in $\mathbb{R}^d$.

Given $\mathbf{y} \in \mathbb{R}^N$, we think of the symbols of $\mathbf{y}$ as residing on the edges of the bipartite graph $G = (V_L, V_R, E)$ (we identify the coordinates of $\mathbb{R}^N$ with the N edges of G in some canonical way). Suppose that for some codeword $\mathbf{c} \in T(G, L)$, $\mathbf{y} = \mathbf{c} + \mathbf{f}$ with $\|\mathbf{f} - \sigma_t(\mathbf{f})\|_1 \leqslant \gamma$. Our goal is to decode $\mathbf{y}$ and recover a close approximation $\mathbf{c}'$ to $\mathbf{c}$, satisfying $\|\mathbf{c}' - \mathbf{c}\|_1 \leqslant \gamma \cdot 2^{O(\log_d n)}$. The vector $\mathbf{c}'$ output by our algorithm will, however, not be guaranteed to belong to $T(G, L)$.

[2]In the double cover operation, we take two copies – $V_L = \{u_L \mid u \in V(H)\}$ and $V_R = \{v_R \mid v \in V(H)\}$ – of the vertex set of H, and connect u_L to v_R whenever (u, v) is an edge of H.

A. Iterative signal decoding algorithm

The algorithm is a natural iterative one, based on a series of "local decoding" steps for the subspace L. We begin with the vector $\mathbf{y}$ residing on the edges of G. For $p = O(\log_d n)$, the algorithm runs p iterations of local decoding of left nodes V_L, alternated with p iterations of decoding at the right nodes V_R. The local decoding at a node $v \in V_L \cup V_R$ is defined as follows:

> if $z_v \in \mathbb{R}^d$ is the local vector residing on $E(v)$, the edges of G incident at v, then run ℓ_1-minimization to find z'_v with minimum $\|z'_v\|_1$ satisfying $Mz'_v = Mz_v$ and replace z_v by $z_v - z'_v$.

At the end of p iterations, the algorithm outputs the current vector residing on the edges of G as the approximation $\mathbf{c}' \in \mathbb{R}^N$ to the codeword $\mathbf{c}$.

B. Analysis of the algorithm

We will use the following notation: For a vector $\mathbf{x} \in \mathbb{R}^E$ and a subset $A \subset E$, $\mathbf{x}_A$ denotes the projection of $\mathbf{x}$ onto the coordinates in A, with the rest of the coordinates zeroed out.

By assumption, $\mathbf{y} = \mathbf{c} + \mathbf{f}$ for some $\mathbf{c} \in T(G, L)$ and $\mathbf{f} \in \mathbb{R}^N$ satisfying $\|\mathbf{f} - \sigma_t(\mathbf{f})\|_1 \leqslant \gamma$. Let $F \subseteq E$, $|F| = t$, be the set of t largest coordinates in absolute value of $\mathbf{f}$. We have $\|\mathbf{f} - \mathbf{f}_F\|_1 \leqslant \gamma$.

For $i = 1, 2, \dots, p$, define the subsets $S_i \subset V_L$ and $T_i \subset V_R$ as follows. S_1 consists of the nodes of V_L that are adjacent to w or more edges in F (where $w = \alpha d$ is as defined in (6)). Clearly, $|S_1| \leqslant |F|/w = \alpha n/2$. For $i \geqslant 1$, define T_i to consist of the nodes of V_R that have w or more neighbors in S_i, and for $i \geqslant 2$, define S_i to consist of the nodes of V_L that have w or more neighbors in T_{i-1}.

By the expander mixing lemma (see for instance [1]), the number of edges $e(S, T)$ between a subset $S \subseteq V_L$ and a subset $T \subseteq V_R$ satisfies $|e(S,T) - \frac{d|S||T|}{n}| \leqslant \lambda\sqrt{|S||T|}$. A simple calculation using the expander mixing lemma implies the following when $\lambda = O(\sqrt{d})$: if $|S_i|/n \leqslant \alpha/2$ (which is satisfied for $i = 1$), $|T_i| \leqslant O(1/(\alpha^2 d))|S_i|$, and if $|T_i|/n \leqslant \alpha/2$, then $|S_{i+1}| \leqslant O(1/(\alpha^2 d))|T_i|$. Hence after $p = O_\alpha(\log_d n)$ iterations, we will have $T_p = \emptyset$.

Let $A_i \subset E$ be the set of edges that are **not** incident on a vertex of S_i, and let $B_i \subset E$ be the set of edges are not incident on a vertex of T_i. Since $T_p = \emptyset$, we have $B_p = E$. Let $\mathbf{a^{(i)}} \in \mathbb{R}^{nd}$ and $\mathbf{b^{(i)}} \in \mathbb{R}^{nd}$ be vectors that result after the i'th iteration of left decoding and right decoding respectively. Note that our final output $\mathbf{c}'$ for the approximation to the true codeword $\mathbf{c}$ is simply $\mathbf{b^{(p)}}$.

It is readily checked that the performance guarantee of the local ℓ_1-minimization decoder for the subspace L implies the following inequalities. Each ℓ_1-minimization based local decoding loses a factor of C in the ℓ_1 error on a particular neighborhood $E(v)$ (as per the guarantee (7)). These neighborhoods being disjoint, we can bound the overall blowup factor in ℓ_1 error per round by C. The first iteration of left decoding gives the guarantee

$$\|\mathbf{a^{(1)}}_{A_1} - \mathbf{c}_{A_1}\|_1 \leqslant C \cdot \gamma \ . \tag{8}$$

For $1 \leqslant i \leqslant p$, the i'th iteration of right decoding implies

$$\|\mathbf{b^{(i)}}_{B_i} - \mathbf{c}_{B_i}\|_1 \leqslant C \cdot \|\mathbf{a^{(i)}}_{A_i} - \mathbf{c}_{A_i}\|_1 \ , \tag{9}$$

and for $1 < i \leqslant p$, the i'th iteration of left decoding implies

$$\|\mathbf{a^{(i)}}_{A_i} - \mathbf{c}_{A_i}\|_1 \leqslant C \cdot \|\mathbf{b^{(i-1)}}_{B_{i-1}} - \mathbf{c}_{B_{i-1}}\|_1 \ . \tag{10}$$

All of the above follow from the fact that if there are less than w "big" errors (compared to the true codeword $\mathbf{c}$) in the local neighborhood $E(v)$ of a node v and the errors (w.r.t $\mathbf{c}_{|E(v)}$) on all but these w coordinates add up to at most ε in ℓ_1-sense, then the local decoding at v gives a codeword of L with ℓ_1-distance at most $C\varepsilon$ from $\mathbf{c}_{|E(v)}$.

Since $B_p = E$, combining the above inequalities, we get

$$\|\mathbf{b^{(p)}} - \mathbf{c}\|_1 \leqslant C^{2p}\gamma \ .$$

Since $p = O(\log_d n)$, the claimed error bound (5) of Theorem 3 follows.

References

[1] N. Alon and J. Spencer. *The Probabilistic Method*. John Wiley and Sons, Inc., 1992.

[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] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. B. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, January 2008.

[4] R. Berinde, P. Indyk, and M. Ruzic. Practical near-optimal sparse recovery in the l_1 norm. In *Proceedings of the 46th Annual Allerton Conference on Communication, Control, and Computing*, 2008.

[5] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[6] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, 51(12):4203–4215, 2005.

[7] 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.

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

[9] V. Guruswami, J. Lee, and A. Razborov. Almost Euclidean sections of ℓ_1^n via expander codes. *Combinatorica*, 2009. To appear.

[10] V. Guruswami, J. Lee, and A. Wigderson. Euclidean sections with sublinear randomness and error-correction over the reals. In *Proceedings of the 12th International Workshop on Randomized Techniques in Computation*, pages 444–454, August 2008.

[11] P. Indyk and M. Ruzic. Near-optimal sparse recovery in the l1 norm. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 199–207, 2008.

[12] 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.

[13] B. S. Kashin and V. N. Temlyakov. A remark on compressed sensing. *Mathematical Notes*, 82(5-6):748–755, 2007.

[14] 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.

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

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

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

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