Uncertainty Principles, Extractors, and Explicit Embeddings of L2 into L1

Piotr Indyk*
MIT
indyk@mit.edu

ABSTRACT

We give an explicit construction of a constant distortion embedding F of l_2^n into l_1^m, with $m = n^{1+o(1)}$. As a bonus, our embedding also has good computational properties: for any input x, Fx can be computed in $n^{1+o(1)}$ time. The previously known mappings required $\Omega(n^2)$ evaluation time.

Categories and Subject Descriptors

F.2.1 [**Analysis of Algorithms and Problem Complexity**]: Numerical Algorithms and Problems

Keywords

Norm embeddings, uncertainty principles, extractors

1. INTRODUCTION

The area of *geometric functional analysis* is concerned with studying the properties of geometric (*normed*) spaces. A typical question in the area is: for two spaces X and Y, equipped with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$, under which conditions is there an *embedding* $F : X \to Y$ such that for any $p, q \in X$, we have $\|p-q\|_X \leq \|F(p)-F(q)\|_Y \leq C\|p-q\|_Y$ for some constant[1] $C \geq 1$?

A prototypical result in the area is the theorem by Dvoretzky [10]: there exists a function $m(n, \epsilon)$, such that the n-dimensional Euclidean space l_2^n can be embedded, with $1+\epsilon$ distortion, into any normed space $M = (X, \|\cdot\|_X)$ of dimension $m(n, \epsilon)$. Many special cases of this theorem, for specific "host spaces" M, have been investigated as well. A notable example is the case of l_1^m, known as the "Dvoretzky's theorem for l_1^m". For this case, it has been shown [11] that there exist embeddings of l_2^n into l_1^m with both the distortion and the "dimension blowup" m/n bounded by a constant; in general, $m(n, \epsilon)$ can be exponential in n.

*This work was supported in part by NSF CAREER grant CCR-0133849 and David and Lucille Packard Fellowship.

[1] C is called the *distortion* of the mapping F.

The method of choice for constructing norm embeddings is the *probabilistic method* (see [29], Section 4, for a very recent overview). Typically, an embedding is chosen at random from some distribution, and one shows that it "works" with high probability. This approach has been introduced by Milman in his novel proof of Dvoretzky's theorem [23] and has been widely used since then, e.g., in the aforementioned paper [11]. Unfortunately, this approach does not lead to a concrete (or explicit) example of an embedding.

The problem of finding an *explicit* embedding of l_2^n into l_1^m has been a subject of several studies (see Figure 1). For constant distortion the best known result, attributed to Rudin [26], guarantees $m = O(n^2)$; see also [18]. The problem of finding an explicit embedding with "low" distortion and dimension blowup has been posed, e.g., in [17] (Section 2.2), or in [22] (Problem 8), or in [21] (Problem 2.1). See also Section 4 of [29] for a related discussion.

In this paper we make progress on this question: for any $1 > \eta > 1/\log n$, we give an explicit [2] construction of an embedding of l_2^n into l_1^m with distortion $1+\eta$, with $m = n2^{O((\log\log n)^2)} = n^{1+o(1)}$. Thus, our result matches the parameters of the non-explicit construction up to a sublinear factor in the dimension bound.

As a bonus, our embedding F has very good *computational* properties. That is, for any vector $x \in \mathbb{R}^n$, the image $Fx \in \mathbb{R}^m$ can be computed in time $n2^{O(\log^2 \log n)}$. This significantly improves over the $O(n^2)$ time previously known, speeding up the algorithms constructed using the embedding (see the next section).

Both explicitness and efficient evaluation stem from the fact that our embedding has a nice *structure*, which can be exploited in multiple ways. Thus, from a high-level perspective, this result enhances our understanding of constructing "structured" embeddings, along the lines of [27, 1, 2][3].

1.1 Applications

As a consequence of our result we obtain, for any constant $\epsilon > 0$, a randomized "Las Vegas" $(1+\epsilon)$-approximate data structure for the *near neighbor* problem in $\mathbb{R}^n$ under the l_2 norm. Our data structure has query time polynomial in the dimension n, and uses space polynomial in the input size. Such result was known for the l_1 norm [15] (refer to

[2] Formally, in this paper, the term "explicit" is defined as "constructible in time that is polynomial in n". See Section 1.2 for an overview of the construction.

[3] In fact, this work has been inspired by the result of [2].

Distortion	Dimension m	Reference	Randomness
$1+\eta$	$O(n)$	[11, 27]	$O(n^2)$
		[16]	$O(n\log^2 n)$
		[3]	$O(n\log n)$
		[20]	$O(n)$
$\sqrt{3}$	$O(n^2)$	[26]	explicit
		(cf. [18])	explicit
$1+1/n$	$n^{O(\log n)}$	[16]	explicit
$1+1/\log n$	$n2^{O((\log\log n)^2)}$	this paper	explicit

Figure 1: Parameters of the known embeddings of l_2^n into l_1^m. The parameter $\eta > 0$ is assumed to be an absolute constant. For probabilistic constructions, the number of required random bits is shown.

that paper for a detailed description of the problem and the results). The extension to the l_2 norm follows immediately from our embedding.

In a similar fashion, our result implies a *deterministic* polynomial-time reduction from several approximate lattice problems under l_2 norm to analogous problems under l_1 norm, with only $1+\epsilon$ loss in the approximation factor. Previously known reductions incurred a loss of a $\sqrt{3}$ factor. The reduction applies to problems such as the *Shortest Vector Problem* and the *Closest Vector Problem*. See [25] for more on this topic.

Very recently, it was shown [19] that an embedding of l_2 into l_1 (in its dual form) can be used to construct *robust vector quantizers*. Our result yields an explicit construction of such quantizers.

1.2 Our techniques

Our embedding is constructed using the recursive "divide and conquer" approach. The main step is provided by the following theorem (see Preliminaries for the notation).

THEOREM 1.1. *For any constants $\zeta, \kappa > 0$, there is an explicit linear mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, $m = n\log^{O(1)} n/\zeta^{O(1)}$, a scaling factor $S > 0$ and an explicit partitioning of the coordinate set $[m]$ into "block" sets $B_1 \ldots B_b$, each of size $n^{1/2+\kappa}2^{(\log\log n)^{O(1)}}/\zeta^{O(1)}$, such that for any $x \in \mathbb{R}^n$, $\|x\|_2 = 1$, we have*

$$(1 - O(\zeta))S \le \sum_{j=1}^{b} \|(Fx)_{B_j}\|_2 \le S$$

By applying this theorem $O(\log\log n)$ times we obtain a mapping $H : \mathbb{R}^n \to \mathbb{R}^{m'}$, $m' = n^{1+o(1)}$, and blocks $B_1 \ldots B_{b'}$ of constant size, such that for any x as above we have

$$\sum_{j=1}^{b'} \|(Hx)_{B_j}\|_2 = S(1 \pm \eta')$$

for some $\eta' > 0$. Since for any $y \in \mathbb{R}^{O(1)}$ we have $\|y\|_2 \le \|y\|_1 \le O(1) \cdot \|y\|_2$, the mapping H preserves the norm of any unit vector x up to a constant factor. [4] The constant-distortion embedding follows in a standard way from the properties of a norm.

[4]Alternatively, one can use the embedding of [16] within each block, to ensure that the distortion is arbitrarily close to 1. This is the approach we use to obtain our main result.

How to construct the mapping F ? We proceed in two steps. In the first step, a vector x is mapped into $Dx \in \mathbb{R}^{m'}$, $m' = \Theta(n)$, such that (i) the mapping D preserves the l_2 norm, and (ii) most of the "mass" $\|Dx\|_2$ of Dx is "spread" over a "large" set S of coordinates. That is, for any "small" set of coordinates S, we have $\|(Dx)_S\|_2 < \delta\|Dx\|_2$ for some constant $1 > \delta > 0$. Such properties are often called *uncertainty principles* [9]. We note that several *probabilistic* constructions of mappings with such properties have been discovered recently, e.g., in [7, 6] and [2]. See Section 4 for more information on this topic.

The key parameter of such mappings is the upper bound on the size of "a small set of coordinates". In particular, if the upper bound was *linear*, then D itself would have the desired embedding properties. Indeed, consider a set S consisting of $s = cm'$ largest (in magnitude) coordinates in Dx, for some constant $c > 0$. From the hypothetical properties of D, it would follow that $\|(Dx)_{[m']-S}\|_2 \ge \sqrt{1-\delta^2}\|x\|_2 = \Omega(\|x\|_2)$. At the same time, for any $i \in [m'] - S$ we would have that $|(Dx)_i|^2 \le \|Dx\|_2^2/s = O(\|x\|_2^2/n)$. Therefore at least $\Omega(n)$ entries $(Dx)_i$ would have absolute values of $\Theta(\|x\|_2/\sqrt{n})$, which would imply that $\|Dx\|_1 = \Omega(\sqrt{n}\|x\|_2)$. At the same time, we have $\|Dx\|_1 \le \sqrt{m'}\|Dx\|_2 = O(\sqrt{n}\|x\|_2)$. Therefore, the mapping D would have constant distortion (after scaling it by a factor of $1/\sqrt{n}$).

Unfortunately, we do not know how to explicitly construct such a mapping D that works for sets S of linear size. However, we do know how to handle sets of size up to $\Theta(\sqrt{n})$. The existence of such mappings, together with an explicit construction (for *some* $\delta > 0$), was discovered by [9]; see Section 4 for further discussion. Here, we require this property for *any* $\delta > 0$. We show that a certain explicit construction of "highly orthogonal" configuration of vectors in $\mathbb{R}^n$ [5, 32] can be used to construct such mapping D.

In the second step, the coordinates of the vector Dx are duplicated, rearranged, and partitioned into b "buckets" (note that these operations preserve the l_2 norm of a vector up to a scaling factor). The goal of this step is to ensure that the resulting vector Fx has the property that, for "most" buckets B_j, we have $\|(Fx)_{B_j}\|^2 \approx \|Fx\|_2/b$. Note that this is possible only if the "mass" $\|Dx\|_2$ is "spread" over more than b coordinates, which is achieved using the aforementioned properties of the mapping D.

The rearrangement is done using an *extractor*, i.e., an expanding bipartite graph, in which the left vertex set is much larger than the right vertex set; see Section 3 for more information on extractors. Many explicit constructions of extractors are known [28]. To use an extractor in our construction, however, we require that the *maximum* degree of any right vertex is at most a constant factor away from the *average* degree of the right vertices. Although this property does not have to hold in general, one can "enforce" it by removing a few edges and vertices of an extractor. This "trimming" procedure is simple and can be performed in time polynomial (in fact, linear) in the size of the graph.

Combining the "spreading" and "rearranging" steps yields the embedding of Theorem 1.1.

2. PRELIMINARIES

In the paper we use n to denote the dimension of the "guest" space l_2. We assume that $n = 2^{2l}$ for an integer $l > 0$; this can be easily achieved by padding extra dimensions with 0's.

We use $[n]$ to denote the set $\{1 \dots n\}$. For any $x \in \mathbb{R}^n$, and $S \subset [n]$, we use x_S to denote the projection of x on S; that is, $(x_1, \dots, x_n)_{\{i_1, \dots, i_s\}} = (x_{i_1}, \dots, x_{i_s})$.

It will be convenient to perform arithmetic operations (such as addition or subtraction) on two vectors x and y even if their dimensions different. Given vectors $x : X \to \mathbb{R}$ and $y : Y \to \mathbb{R}$, and an operation $\otimes$, the value of $x \otimes y$ is equal to $x' \otimes y'$, where $x' : X \cup Y \to \mathbb{R}$ is an extension of x obtained by padding the dimensions in $Y - X$ with 0's; y' is defined analogously.

In a graph G with the set of edges E, we use $\Gamma_G(i)$ to denote the sequence of vertices adjacent to the vertex i in G (in any fixed order).

Consider any set $S \subset \Sigma^n$ for some set Σ. We say that the uniform distribution over S (or just S) is *k-wise independent*, if for any $|K| \leq k$, $z \in \Sigma^k$, $\Pr_{x \in S}[x_K = z] = |\Sigma|^{-k}$.

3. EXTRACTORS

Our construction utilizes *extractors*. They are defined as follows.

Definition 3.1. *A bipartite graph $G = (A, B, E)$, $A = [a], B = [b]$, with each left node having degree d, is called an (ϵ, l)-*extractor*, if it satisfies the following property. Consider any distribution $\mathcal{P}$ over A such that for any $i \in A$, $\Pr_{\mathcal{P}}[i] \leq 1/l$. Let j be a random variable over B obtained by choosing a vertex i from A, choosing t uniformly at random from $[d]$, and setting $j = \Gamma_G(i)_t$. Let $G(\mathcal{P})$ be the resulting distribution of the random variable j, and let $\mathcal{I}$ be the uniform distribution over B. We require that $G(\mathcal{P})$ and I are* ϵ-close, *i.e., that $\|G(\mathcal{P}) - \mathcal{I}\|_1 \leq \epsilon$, where both distributions are interpreted as vectors in $\mathbb{R}^b$.*

It is useful to observe that if $\mathcal{P}$ is a uniform distribution over some set $A' \subset A$, then, for any $j \in B$, the value of $\Pr_{G(\mathcal{P})}[j]$ is proportional to the number of neighbors of j that belong to A'.

There exist several explicit constructions of extractors which guarantee the left degree $d = 2^{(\log \log a)^{O(1)}}/\epsilon^{O(1)}$ (see [28] for an overview). Here we use the construction of [33], with parameters $l = a^\lambda$, $b = a^{\lambda - \kappa}$ and $d = (\log a)^{O(1)}$, for any constants $\lambda, \kappa \in (0, 1)$. For somewhat simpler extractor constructions (which yield slightly worse parameters of our embedding) see, e.g., [24, 13].

In our applications we also need an upper bound on the right degree Δ of the extractor. To this end we show[5] that one can achieve $\Delta = O(ad/b)$ by taking any extractor G and applying the following "trimming" procedure. Let $avg = ad/b$ be the average right degree of G, and let $t > 1$, $\delta > 0$ be constants determined later.

1. Select $B' \subset B$ containing nodes with degree higher than $t \cdot avg$.
2. Select $A' \subset A$ containing nodes with more than δd neighbors in B'.
3. Construct an induced subgraph $G' = (A - A', B - B', E')$ of G.
4. For any node $i \in A - A'$ with $d_i = |\Gamma_{G'}(i)| < d$, add $d - d_i$ edges incident to i. Let E'' be the set of all of those edges. The right endpoints of the edges are chosen such that each right vertex is incident to at most $\lceil |E''|/|B - B'| \rceil$ of the edges. For example, we can set the s-th edge of E'' to be incident to the node $(s \mod |B - B'|) + 1$.

From the construction it follows that each left node of G' has degree d, and that the right degree of G' is $O(ad/b + ad/|B - B'|)$. The rest is shown in the following claim.

Claim 3.2. *Set $t = 2$ and $\delta = 4\epsilon$. Then the graph G' has $a/2$ left vertices, $b/2$ right vertices, and is an $(\delta + \epsilon, l)$-extractor.*

Proof. First, observe that at most b/t nodes in B have degree higher than $t \cdot avg$, so $|B - B'| \geq b/2$. Let F be the set of edges incident to B', and let $\mathcal{P}$ be a uniform distribution over A. In this case $\frac{|F|}{|E|} = \sum_{j \in B'} \Pr_{G(\mathcal{P})}[j]$. Moreover, the probability of choosing $j \in B'$ with respect to distribution $G(\mathcal{P})$ is at least t/b, which exceeds the average $1/b$ by at least $(t-1)/b$. Since

$$\sum_{j \in B'} \Pr_{G(\mathcal{P})}[j](1 - 1/t) \leq \sum_{j \in B'} \Pr_{G(\mathcal{P})}[j] - 1/b \leq \epsilon$$

it follows that $\frac{|F|}{|E|}(1 - 1/t) \leq \epsilon$, or $\frac{|F|}{|E|} \leq \epsilon(1 + \frac{1}{t-1}) = 2\epsilon$. Therefore, at most $2\epsilon/\delta = 1/2$ fraction of nodes in A have more than δd edges leading to B', so $|A - A'| \geq a/2$. Finally, each node in $A - A'$ is incident to at most δd new edges E''. Thus, for any distribution $\mathcal{P}$ over A', we have $\|G(\mathcal{P}) - G'(\mathcal{P})\|_1 \leq \delta$. It follows that G' is an $(\delta + \epsilon, l)$-extractor.

□

4. UNCERTAINTY PRINCIPLES

Let H be a normalized $n \times n$ Hadamard matrix. That is, all entries of H are either $-1/\sqrt{n}$ or $1/\sqrt{n}$, and the rows of H are mutually orthogonal. Thus, the rows of H form an orthonormal basis for $\mathbb{R}^n$. In addition, let I be the $n \times n$ identity matrix.

The pair of matrices (I, H) has the following nice property: for each row u of I and row v of H, we have $|u \cdot v| \leq 1/\sqrt{n}$. In general, consider any collection $H_1 \cup \ldots \cup H_L$ of $n \times n$ orthonormal matrices, and let D be an $n \cdot L \times n$ matrix obtained by concatenating the rows of those matrices (D is often called a *dictionary*). The *coherence* of D is defined as the maximum, over all rows u, v of D, of $|u \cdot v|$. In particular, the coherence of the concatenation of matrices I and H has coherence $1/\sqrt{n}$, while the coherence of any orthonormal matrix is 0.

In [9] the authors show a relation between the coherence of D, and the sparsity of the representation Dx of any vector $x \in \mathbb{R}^n$. Specifically, they consider a dictionary D obtained from the identity matrix I and a Fourier matrix; such a dictionary D has coherence $1/\sqrt{n}$. They show that, for any non-zero $x \in \mathbb{R}^n$, Dx must have at least $2\sqrt{n}$ non-zero elements. They also show a more general fact: there exists $c > 0$, such that for any set S of coordinates of Dx of size at most $c\sqrt{n}$, the ratio $\frac{\|(Dx)_S\|_2}{\|Dx\|_2}$ is bounded away from 1 (although not arbitrarily close to 0).

[5] This fact was apparently known, but we were unable to locate an appropriate reference. However, the paper [31] provides such a proof for somewhat simpler combinatorial objects called *dispersers*.

This phenomenon is akin to an *uncertainty principle*: a signal cannot be concentrated in both the time and the frequency domain.

This result was generalized [8] to dictionaries D consisting of two arbitrary orthonormal matrices H_1, H_2. They show that, for any non-zero $x \in \mathbb{R}^n$, the vector Dx must have at least $1/M$ non-zero entries, where M is the coherence of D. This theorem was later extended in [12] to concatenations of L orthonormal matrices $H_1 \dots H_L$.

In this paper, we exploit a general version of the uncertainty principle phenomenon, stated as Lemma 4.2 below. We start from the following observation.

CLAIM 4.1. *Consider any $N \times n$ dictionary D with coherence M, and a submatrix D_S consisting of at most s rows in D. For any unit vector $x \in \mathbb{R}^n$, we have $\|D_S x\|_2^2 \le 1 + Ms$.*

PROOF. We need to upper-bound the 2-norm $\|D_S\|_2$ of D_S, which is equal to the square root of the largest eigenvalue $\lambda(G)$ of the Gram matrix $G = D_S \times D_S^T$. Clearly, for $i, j \in [s]$, $i \ne j$, we have $G_{ii} = 1$ and $|G_{ij}| \le M$. It follows that $|\lambda(G)| \le |\lambda(I)| + |\lambda(G - I)| \le 1 + Ms$, since all entries of $G - I$ are in the range $[-M, M]$. □

Note: a related proof has appeared in [30] (Proposition 4.3).

Assume now that D is obtained by concatenating L orthonormal matrices $H_1 \dots H_L$. Clearly, for any unit vector $x \in \mathbb{R}^n$, we have $\|Dx\|_2^2 = L$. At the same time, for any set of coordinates $S \subset [Ln]$, $|S| = s$, we have $\|(Dx)_S\|_2^2 \le 1 + Ms$ (Claim 4.1). This shows the following lemma.

LEMMA 4.2. *Let D be a dictionary obtained by concatenating rows of L orthonormal $n \times n$ matrices $H_1 \dots H_L$ with coherence M. Then, for any set of coordinates $S \subset [Ln]$, $|S| = s$, and any unit vector $x \in \mathbb{R}^n$, we have*

$$\|(Dx)_S\|_2^2 \le (1 + Ms)\|Dx\|_2^2/L$$

It remains to construct dictionaries with small coherence. Fortunately, if $n = 2^{2k}$ for an integer $k > 0$, then there exist explicit constructions of a dictionary D consisting of the rows of $L = n/2 + 1$ orthonormal matrices $H_1 \dots H_L$, such that the coherence of D is $1/\sqrt{n}$. The constructions appear in [5, 32, 4]; see [14] for a simplified description of the construction (In a nutshell: each H_i is either an identity matrix, or is equal to $H \times D_i$, where H is the Hadamard matrix, and D_i's are carefully chosen diagonal matrices.) Therefore, for $s = \sqrt{n}$, we have $Ms \le 1$, and any s coordinates of Dx can contain at most $\sqrt{2/L}$ fraction of the "mass" $\|Dx\|_2$.

COROLLARY 4.3. *Let $n = 2^{2k}$ for an integer $k > 0$. For any $\delta > 2/\sqrt{n}$, there exists an explicit construction of a dictionary D consisting of the rows of $L = 2/\delta^2$ orthonormal $n \times n$ matrices $H_1 \dots H_L$, such that for any unit vector $x \in \mathbb{R}^n$, and any set $S \subset [nL]$, $|S| \le s = \sqrt{n}$, we have $\|(Dx)_S\|_2 \le \delta\|Dx\|_2$. In addition, given $x \in R^n$, the vector Dx can be computed in $O(nL \log n)$ time using Fast Hadamard Transform.*

5. THE CONSTRUCTION

Our construction utilizes the observation of Corollary 4.3. Fix $\delta > 0$ to be some constant, and let $L = O(1/\delta^2)$ and $s = \sqrt{n}$ be as in the latter corollary. In addition, we use an (ϵ, l)-extractor G with $A = [Ln]$ and $B = [b]$ for $b = n^{1/2-\kappa}$, $\kappa > 0$, $l = (1-\delta)^2 s/L$, left degree $d = (\log a)^{O(1)}$ and right degree $\Delta = O(nLd/b)$. The values of δ and ϵ are set at the end of the following proof of the the Theorem 1.1. For convenience, we restate the theorem below, using the notation introduced above.

Theorem 1.1. For any $\zeta, \kappa > 0$, there is an explicit linear mapping $F : \mathbb{R}^n \to \mathbb{R}^m$, $m = O(nLd) = n \log^{O(1)} n/\zeta^{O(1)}$, and a partitioning of the coordinates set $[m]$ into sets $B_1 \dots B_b$, each of size $O(\Delta) = O(nLd/b) = n^{1/2+\kappa} 2^{(\log\log n)^{O(1)}}/\zeta^{O(1)}$, such that for any $x \in \mathbb{R}^n$, $\|x\|_2 = 1$, we have

$$(1 - O(\zeta))\sqrt{Ldb} \le \sum_{j=1}^{b} \|(Fx)_{B_j}\|_2 \le \sqrt{Ldb}$$

PROOF. The mapping F is constructed as

$$Fx = \oplus_{j=1}^{b} (Dx)_{\Gamma_G(j)}$$

where $\oplus$ is the *direct sum* operator (i.e., concatenation).

Let $y = Dx$. Observe that $\|y\|_2 = \sqrt{L}$. In the following we show that the mapping has constant expansion as well as constant contraction, with respect to the norm of the host space.

Expansion. Let $y^d = \oplus_{i=1}^{d} y$. Note that $\|y^d\|_2 = \sqrt{Ld}$. We can interpret Fx as a vector y^d partitioned into d blocks, and use the following lemma.

LEMMA 5.1. *For any $x \in \mathbb{R}^n$, and any partitioning $P_1 \dots P_b$ of $[n]$, we have*

$$\sum_{t=1}^{b} \|x_{P_t}\|_2 \le \sqrt{b}\|x\|_2$$

PROOF. Construct a vector $y \in \mathbb{R}^b$ such that $y_j = \|x_{P_t}\|_2$. Observe that $\|y\|_1 = \sum_{t=1}^{b} \|x_{P_t}\|_2$ and $\|x\|_2^2 = \sum_{t=1}^{b} \|x_{P_t}\|_2^2 = \sum_{t=1}^{b} y_t^2 = \|y\|_2^2$. Since $\|y\|_1 \le \sqrt{b}\|y\|_2$, the lemma follows. □

From Lemma 5.1 it follows that

$$\sum_{i=1}^{b} \|y_{\Gamma_G(i)}\|_2 \le \sqrt{b}\|y^d\|_2 = \sqrt{Ldb}$$

Contraction. Let $S \subset [Ln]$ be the set of indices of the s largest (in magnitude) entries of y. Consider the vector $z = y_{[Ln]-S}$ interpreted as a vector in $\mathbb{R}^{Ln}$. Let $\rho = \|z\|_2$. From Lemma 4.3 we know that $\rho \ge \sqrt{L}(1-\delta)$. At the same time, for each $i = 1 \dots Ln$, we have $z_i^2 \le \|y\|_2^2/s = L/s$.

We use z to construct a probability distribution $\mathcal{P}$ over $[Ln]$, by defining $p_i = z_i^2/\|z\|_2^2 = z_i^2/\rho^2$. It follows that $p_i \le \frac{L/s}{\rho^2} \le 1/l$. Therefore, $\mathcal{P}$ satisfies the conditions on using the extractor G. This implies that the distribution $\mathcal{Q} = G(\mathcal{P})$ over B is ϵ-close to the uniform distribution over B.

For any $j \in B$, the probability q_j with respect to $\mathcal{Q}$ is equal to

$$q_j = 1/d \cdot \sum_{i \in \Gamma_G(j)} p_i = \frac{1}{\rho^2 d} \cdot \sum_{i \in \Gamma_G(j)} z_i^2 = \frac{1}{\rho^2 d}\|z_{\Gamma_G(j)}\|_2^2$$

Since $\mathcal{Q}$ is ϵ-close to the uniform distribution, we get

$$\sum_{j=1}^{b} \left| \frac{1}{\rho^2 d} \|z_{\Gamma_G(j)}\|_2^2 - 1/b \right| \leq \epsilon$$

Therefore, for at least $b(1-\alpha)$ indices j, we have $\frac{1}{\rho^2 d}\|z_{\Gamma_G(j)}\|_2^2 \geq (1-\epsilon/\alpha)/b$, or, $\|z_{\Gamma_G(j)}\|_2 \geq \sqrt{(1-\epsilon/\alpha)\rho^2}\sqrt{d/b}$. Therefore

$$\sum_j \|y_{\Gamma_G(j)}\|_2 \geq \sum_j \|z_{\Gamma_G(j)}\|_2 \geq b(1-\alpha)\sqrt{\rho^2(1-\epsilon/\alpha)}\sqrt{d/b}$$

$$\geq (1-\alpha)\sqrt{(1-\delta)^2(1-\epsilon/\alpha)} \cdot \sqrt{Ldb}$$

Setting $\alpha = \delta = \zeta$, and $\epsilon = \zeta^2$ finishes the proof.

COROLLARY 5.2. *For any $1 > \eta > 0$ there is an explicit embedding F of l_2^n into l_1^m,*

$$m = (\log^{O(1)}(n)/\eta^{O(1)})^{O(\log\log n)} \cdot 1/\eta^{O(\log 1/\eta)}$$

with distortion $1 + O(\eta)$.

PROOF. Apply Theorem 1.1 $O(\log\log n)$ times with $\zeta = \eta/\log\log n$ to reduce each block size to a constant $n' = \Theta(1/\eta)$. Within each block use the embedding of [16] which embeds $l_2^{n'}$ into $l_1^{n'^{O(\log n')}}$ with distortion $1+\eta$. The total distortion is at most $(1+\eta)(1+\eta/\log\log n)^{O(\log\log n)} = 1 + O(\eta)$. The dimension blowup is at most

$$(\log^{O(1)}(n)/\zeta^{O(1)})^{O(\log\log n)} \cdot 1/\eta^{O(\log 1/\eta)}$$

Note that the latter bound is at most $2^{O((\log\log n)^2)}$ for $\eta = 1/\log n$. □

REMARK 5.3. *Consider the case where $\eta > 1/\log n$. In this case, given $x \in R^n$, the vector Fx can be computed in $2^{O(\log^2 \log n)}$ time. This follows from Corollary 4.3 and the recursive construction of the embedding.*

Acknowledgments

The author would like to thank Assaf Naor, Gideon Schechtman, Martin Strauss, Tasos Sidiropoulos, Chris Umans, Salil Vadhan and David Woodruff, for many helpful discussions and/or comments on early drafts of this paper.

6. REFERENCES

[1] D. Achlioptas. Database-friendly random projections. *Proceedings of the Symposium on Principles of Database Systems*, pages 274–281, 2001.

[2] N. Ailon and B. Chazelle. Approximate nearest neighbors and the Fast Johnson-Lindenstrauss Transform. *Proceedings of the Symposium on Theory of Computing*, 2006.

[3] S. Artstein-Avidan and V. D. Milman. Logarithmic reduction of the level of randomness in some probabilistic geometric constructions. *Journal of Functional Analysis*, 235:297–329, 2006.

[4] A. R. Calderbank, P. J. Cameron, W. M . Kantor, and J. J. Seidel. Z4-Kerdock codes, orthogonal spreads and extremal euclidean line-sets. *Proceedings of the London Mathematical Society*, 2000.

[5] P. J. Cameron and J.J. Seidel. Quadratic forms over GF(2). *Indag. Math*, 1973.

[6] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52:489 – 509, 2006.

[7] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289 – 1306, 2006.

[8] D. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *Transactions on Information Theory*, 47:2845–2862, 2001.

[9] D. Donoho and P. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 1989.

[10] A. Dvoretzky. A theorem on convex bodies and applications to banach spaces. *Proceedings of the National Academy of Sciences USA*, 45:223–226, 1959.

[11] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Mathematica*, 139:53–94, 1977.

[12] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 2003.

[13] V. Guruswami, C. Umans, and S. Vadhan. Extractors and condensers from univariate polynomials. *IEEE Conference on Communication Complexity*, 2007.

[14] R. W. Heath, T. Strohmer, and A. J. Paulraj. On quasi-orthogonal signatures for CDMA systems. *IEEE Transactions on Information Theory*, 52:1217–1226, 2006.

[15] P. Indyk. Dimensionality reduction techniques for proximity problems. *Proceedings of the Ninth ACM-SIAM Symposium on Discrete Algorithms*, 2000.

[16] P. Indyk. Stable distributions, pseudorandom generators, embeddings and data stream computation. *Proceedings of the Symposium on Foundations of Computer Science*, 2000.

[17] W.B. Johnson and G. Schechtman. Finite dimensional subspaces of l_p. In W.B. Johnson and J. Lindenstrauss, editors, *Handbook of the Geometry of Banach Spaces*, pages 837–870. Elsevier, 2001.

[18] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Proceedings of 35th Annual IEEE Symposium on Foundations of Computer Science*, pages 577–591, 1994.

[19] Yu. Liubarskii and R. Vershynin. Uncertainty principles and vector quantization. *Manuscript*, 2006.

[20] S. Lovett and S. Sodin. Almost euclidean sections of the n-dimensional cross-polytope using o(n) random bits. *ECCC Report TR07-012*, 2007.

[21] J. Matoušek. Collection of open problems on low-distortion embeddings of finite metric spaces. *Available at* `http://kam.mff.cuni.cz/~matousek/metrop.ps.gz`, 2004.

[22] V. Milman. Topics in asymptotic geometric analysis. *GAFA - Geometric and Functional Analysis*, pages 792–815, 2000.

[23] V.D. Milman. A new proof of the theorem of A. Dvoretzky on sections of convex bodies. *Funct. Anal. Appl*, 1971.

[24] R. Raz, O. Reingold, and S. Vadhan. Extracting all the randomness and reducing the error in trevisan's extractors. *Proceedings of the Symposium on Theory of Computing*, 1999.

[25] O. Regev and R. Rosen. Lattice problems and norm embeddings. *Proceedings of the Symposium on Theory of Computing*, 2006.

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

[27] G. Schechtman. Random embeddings of euclidean spaces in sequence spaces. *Israel J. Math.*, 40:187–192, 1981.

[28] R. Shaltiel. Recent developments in extractors. In *Current trends in theoretical computer science. The Challenge of the New Century. Vol 1: Algorithms and Complexity*, 2004.

[29] S. Szarek. Convexity, complexity and high dimensions. *Proceedings of the International Congress of Mathematicians (manuscript available from the author's home page)*, 2006.

[30] J. Tropp. Topics in sparse approximation. *Ph.D. dissertation, Computational and Applied Mathematics, UT-Austin*, 2004.

[31] A. Wigderson and D. Zuckerman. Expanders that beat the eigenvalue bound: Explicit constructions and applications. *Combinatorica*, 1999.

[32] W.K. Wooters and B. D. Fields. Optimal state-determination by mutually unbiased measurements. *Annals of Physics*, 191:363–381, 1989.

[33] D. Zuckerman. Randomness-optimal oblivious sampling. *Random Structures and Algorithms*, 11:345–367, 1997.