Electronic Colloquium on Computational Complexity, Report No. 162 (2010)

Deterministic Construction of a high dimensional ℓ_p section in ℓ_1^n for any $p < 2$

Zohar S. Karnin[**]

Abstract

For any $0 < r < p < 2$, and $\epsilon > 0$, we give an efficient deterministic construction of a linear subspace $V \subseteq \mathbb{R}^n$, of dimension $(1-\epsilon)n$ in which the ℓ_p and ℓ_r norms are the same up to a multiplicative factor of $\mathrm{poly}(\epsilon^{-1})$ (after the correct normalization). As a corollary we get a deterministic compressed sensing algorithm (Base Pursuit) for a new range of parameters. In particular, for any constant $\epsilon > 0$ and $p < 2$, we obtain a linear operator $A : \mathbb{R}^n \to \mathbb{R}^{\epsilon n}$ with the ℓ_1/ℓ_p guarantee for $(n \cdot \mathrm{poly}(\epsilon))$-sparse vectors. Namely, let x be a vector in $\mathbb{R}^n$ whose ℓ_1 distance from a k-sparse vector (for some $k = n \cdot \mathrm{poly}(\epsilon)$) is δ. The algorithm, given Ax as input, outputs an n dimensional vector y such that $||x-y||_p \leq \delta k^{1/p-1}$. In particular this gives a weak form of the ℓ_2/ℓ_1 guarantee.

Our construction has the additional benefit that when viewed as a matrix, A has at most $O(1)$ non-zero entries in each row. As a result, both the encoding (computing Ax) and decoding (retrieving x from Ax) can be computed efficiently.

[*]Faculty of Computer Science, Technion, Haifa 32000, Israel. Email: zkarnin@cs.technion.ac.il.

ISSN 1433-8092

1 Introduction

A typical question, related to dimensional reduction, in geometric functional analysis is: For two spaces X and Y with norms $||\cdot||_X$ and $||\cdot||_Y$, when does there exist an embedding $F : X \to Y$ with the property that for every $\alpha \in X$, we have $||\alpha||_X \leq ||F(\alpha)||_Y \leq C||\alpha||_X$ for some constant $C > 1$ (we refer to C as the distortion of the embedding F)?

A theorem by Dvoretzky [Dvo59] asserts that there exists a function $n(m, \epsilon)$, such that for any normed space Y of dimension $n(m, \epsilon)$, the m dimensional Euclidian space ℓ_2^m can be embedded, with distortion $C = 1 + \epsilon$, into Y. A special case of interest to us is when $Y = \mathbb{R}^n$ equipped with the ℓ_1 norm.

The image of the embedding $V = F(\ell_2^m) \subseteq \mathbb{R}^{n(m,\epsilon)}$ is called an Euclidian section in $\ell_1^{n(m,\epsilon)}$. We use the notion of distortion in the same manner w.r.t. both an embedding or a section. That is, the distortion of a subspace $V \subseteq \ell_1^n$ of $\dim(V) = m$ is defined as the distortion of an embedding F mapping ℓ_2^m into V. In Dvoretzky's theorem, the expression $n(m, \epsilon)$ is non-optimal for the case of the ℓ_1 norm. Moreover, the proof is non-constructive in the sense that it only guarantees the existence of such a section and does not provide an algorithm for finding one. Classical results in high dimensional geometry [FLM77, Kas77] assert that for a constant value of ϵ, a random subspace $X \subseteq \mathbb{R}^n$ of dimension δn [FLM77] or even $(1 - \delta)n$ [Kas77] is an euclidian section in ℓ_1^n w.h.p.. Namely, for any ϵ there exist a subspace of dimension $c_\epsilon n$ with distortion $1 + \epsilon$. Alternatively, there also exists a subspace of dimension $(1 - \epsilon)n$ and distortion C_ϵ.

Such subspaces are instrumental in several algorithms for important theoretical and practical problems such as high dimensional nearest neighbor search [Ind06], and compressed sensing [DeV07]. One expects that an explicit construction will lead to a better understanding of the underlying geometric structures and as a result, to improved algorithms for these problems. Consequently, the problem of finding such an explicit Euclidian section has been given much attention over the last years.

In the constructions of [Kas77, FLM77], the number of random bits required for picking the random subspace is $O(n^2 \log(n))$. One direction of research consists of (partially) derandomizing these constructions. Namely, finding a low distortion subspace using fewer random bits. The second research direction is to look for a deterministic construction where we either allow the subspace to be of smaller dimension (i.e., sublinear) or the distortion to be large (i.e., super-constant).

In [AAM05], the authors gave a randomized construction that achieves the same parameters as [Kas77] while using only $O(n \log(n))$ random bits. [LS08] showed how to reduce the number of random bits to $O(n)$ using similar techniques to those of [AAM05]. [GLW08] achieve a constant distortion subspace, of linear dimension using $O(n^\delta)$ random bits (for any constant δ) by applying ideas from coding theory. Recently, [IS10] achieved a subspace with arbitrarily (constant) low distortion with only $O(n^\delta)$ random bits using different methods.

As for explicit constructions, Rudin [Rud60] and later Linial, London and Rabinovich, [LLR95], gave an example of a subspace X with a constant distortion and with $\dim(X) = O(\sqrt{n})$. Indyk [Ind06, Ind07] presented a subspace X with distortion $1 + o(1)$ and $\dim(X) \geq \frac{n}{\exp((\log\log n)^2)}$. Note that these constructions give constant distortion but the dimension is sublinear. In [DS89] a subspace X of $\dim(X) = n/2$ and distortion $O(n^{1/4})$ is (implicitly) constructed. [GLR08] gave a construction of a subspace of dimension $\dim(X) = (1 - \epsilon)n$ and distortion $(\epsilon^{-1} \log\log n)^{O(\log\log n)}$. Table 1 contains a summary of the results.

In this paper we consider a relaxation of the problem at hand. Instead of considering the ℓ_1/ℓ_2 distortion, we consider the ℓ_1/ℓ_p distortion for any $p < 2$. That is, we find a high dimensional subspace in which the ratio between the ℓ_1 and ℓ_p norms is the same in every vector up to some constant multiplicative factor.

Table 1: Parameters of known Euclidean sections in ℓ_1^n

Distortion	Dimension	Randomness	Paper
$O_\epsilon(1)$	$(1-\epsilon)n$	$O(n\log(n))$	[AAM05]
$O_\epsilon(1)$	$(1-\epsilon)n$	$O(n)$	[LS08]
$O_{\epsilon,\delta}(1)$	$(1-\epsilon)n$	$O(n^\delta)$	[GLW08]
$1+\epsilon$	$(\delta\epsilon)^{O(\delta^{-1})}n$	$O(n^\delta)$	[IS10]
$1+1/n$	$2^{O(\sqrt{\log n})}$	explicit	[Ind06]
$1+1/\log n$	$n2^{-O((\log\log n)^2)} = n^{1-o(1)}$	explicit	[Ind07]
$(\epsilon^{-1}\log\log n)^{O(\log\log n)}$	$(1-\epsilon)n$	explicit	[GLR08]

1.1 Some definitions and statement of our results

We begin with a formal definition of the distortion of a subspace. We give a slightly more general definition than that of the distortion w.r.t. the ℓ_1 and ℓ_2 norms.

Definition 1.1. *Let $0 < r < p$ and let $V \subseteq \mathbb{R}^n$. Denote*

$$C_{\min} = \min_{0\neq x\in V} \frac{||x||_p}{||x||_r} \quad C_{\max} = \max_{0\neq x\in V} \frac{||x||_p}{||x||_r}.$$

The distortion of V w.r.t. ℓ_r and ℓ_p, denoted by $\Delta_{r\to p}(V)$, is defined as $C_{\max}/C_{\min}$

In this paper we consider a new approach towards finding high dimensional subspaces with constant distortion. We deterministically construct a subspace $V \subseteq \mathbb{R}^n$, for any n, in which the distortion between the ℓ_1 and ℓ_p norms is constant, for any $p < 2$ (recall that $||x||_p = (\sum |x_i|^p)^{1/p}$). An immediate corollary is the following: Given a subspace of constant dimension in which ℓ_2 and $\ell_{2-\epsilon}$ are the same (up to a constant multiplicative factor), one can deterministically construct a subspace of almost the same dimension with a constant distortion (w.r.t. ℓ_1 and ℓ_2 norms). The following theorem states our main result.

Theorem 1.2. *For any $\epsilon > 0$ and sufficiently large $n \in \mathbb{N}$, there exists a deterministic algorithm which constructs, in $\mathrm{poly}(n)\cdot\exp(\mathrm{poly}(\epsilon^{-1}))$ time, a matrix A of dimension $\epsilon n\times n$ with the following properties: For any $0 < r < p < 2$, it holds that $\Delta_{r\to p}(\mathrm{Ker}(A)) \le \epsilon^{O(\frac{1}{rp(2-p)})}$ where $\mathrm{Ker}(A)$ is the kernel of A when viewed as a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^{\epsilon n}$.*

This result has two immediate corollaries regarding "proper" Euclidean sections. First, any high dimension Euclidean section in ℓ_p^n for $p < 2$ would lead to an Euclidean section in ℓ_1^n with roughly the same distortion (up to some constant multiplicative factor). Second, by setting $p = 2 - (\log(n)\cdot\log(\epsilon^{-1}))^{-1/2}$, $r = 1$, one can get a section of dimension $(1-\epsilon)n$ and with distortion $n^{o(1)}$ for any constant $\epsilon > 0$. We note however that this result is weaker than that of [GLR08].

Corollary 1.3. *Let $\epsilon > 0$, $n \in \mathbb{N}$ and $A \in \mathbb{R}^{n\times\epsilon n}$ as in Theorem 1.2. Then $\mathrm{Ker}(A)$ is a subspace with dimension $\dim(\mathrm{Ker}(A)) \ge (1-\epsilon)n$ and*

$$\Delta_{1\to 2}(\mathrm{Ker}(A)) \le 2^{O\left(\sqrt{\log(n)\cdot\log(\epsilon^{-1})}\right)}.$$

In particular, for any constant $\epsilon > 0$, $\Delta_{1\to 2}(\mathrm{Ker}(A)) = n^{o(1)}$.

1.2 Application to Compressed Sensing

1.2.1 Background

In the area of Compressed Sensing, given parameters $k, n \in \mathbb{N}$ where $k < n$, the objective is to construct a linear operator $A : \mathbb{R}^n \to \mathbb{R}^{m(n,k)}$ and a matching recovery algorithm with the following properties: For any k-sparse [1] (or "almost k-sparse") vector $x \in \mathbb{R}^n$, given Ax as input, the recovery algorithm must efficiently reconstruct x (or an "approximation" of x). This problem is also known as the problem of sparse signal recovery.

The three major properties of an operator $A : \mathbb{R}^n \to \mathbb{R}^{m(n,k)}$ and its recovery algorithm are the following: The encoding/decoding time, the description length $m(n, k)$, and the approximation guarantee. The encoding time is the running time of the algorithm computing Ax given x. The decoding time is the time required to recover x from Ax. The description length is m, the length of the compressed vector, which we would like to minimize. The approximation guarantee concerns the quality of recovery in the case of *almost sparse* vectors. To define it we introduce some notations. Let $x \in \mathbb{R}^n$ be some vector and let z be the closest k-sparse vector to x. Denote by y the output of the recovery algorithm given Ax as input. We would like the algorithm to satisfy that the distance of y from x is somehow bounded by the distance of z from x. We say that the algorithm has the ℓ_1/ℓ_1 guarantee when $||y - x||_1 = O(||z - x||_1)$. Similarly, it has the ℓ_1/ℓ_2 guarantee when $||y - x||_2 = O(||z - x||_1/\sqrt{k})$. The ℓ_1/ℓ_2 guarantee is a stronger requirement than the ℓ_1/ℓ_1 guarantee as any algorithm with the ℓ_1/ℓ_2 guarantee also has the ℓ_1/ℓ_1 guarantee. For a more thorough review we refer the reader to, e.g., [CT06, GI10, Bar07].

We focus on the case where the sparsity guarantee k is linear. Namely, $k = \Omega(n)$. In this setting all known algorithms achieving an approximation guarantee better than the ℓ_1/ℓ_1 guarantee have two main disadvantages: First, the construction of the encoding matrix A is randomized. Furthermore, there is no known deterministic algorithm that can verify whether a specific matrix A is a good encoding matrix. The second disadvantage is the running time of the encoder and of the decoder. The matrices are usually highly dense and require an encoding time of[2] $\Theta(n^2)$. The sparsest matrix so far was given by [GLW08]: For any $\delta > 0$ they construct a matrix with $\Theta(n^\delta)$ non-zeros in each row leading to an encoding time of $\Theta(n^{1+\delta})$ (the description length grows as δ approaches 0). We note that for the range of parameters that we discussed above (and that is studied in this paper), the decoding is done via a linear program (base pursuit), whose running time is also dependent on the density of A (each row of A is a constraint and thus its density is translated into that of the constraints).

Our result gives an encoding matrix A which has at most $O(1)$ non-zero entries in each row. Given Ax as input, the "base pursuit" algorithm (details are given in the following section) has the ℓ_1/ℓ_p guarantee for any $p < 2$. Namely, using the above notations, $||y-x||_p = O(||z-x||_1 \cdot k^{1/p-1})$. As the matrix is highly sparse, the encoding and decoding (that is done using the mentioned linear program) are more efficient than in previous works. Another advantage of our result is that our construction is deterministic whereas all previous works, for this range of parameters, were randomized.

1.2.2 The application

In [KT07], an explicit connection between Euclidean sections and compressed sensing is given. Assume a linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$ has the property that the distortion between ℓ_2 and ℓ_1 in

[1] A k-sparse vector is a vector with at most k non-zero entries.

[2] We note that for values of $k < n/\text{polylog}(n)$ there are constructions of matrices for which the encoder runs in quasilinear time.

Ker(A) is bounded by D. That is, $\Delta_{1\to 2}(\mathrm{Ker}(A)) \leq D$. Then any vector $x \in \mathbb{R}^n$ with at most $k = n/4D^2$ non-zero entries can be efficiently recovered from Ax via the "Base Pursuit" algorithm which consists of solving the following linear program[3]

$$\min ||y||_1 \text{ subject to } Ay = Ax.$$

[KT07] also prove that the Base Pursuit algorithm has the ℓ_1/ℓ_2 guarantee (when $\Delta_{1\to 2}(\mathrm{Ker}(A)) \leq D$). Via a slight modification of the proof of [KT07], we show that for any $p < 2$, if $\Delta_{1\to p}(\mathrm{Ker}(A)) \leq D$ then A has the ℓ_1/ℓ_p guarantee for $k = \Omega(nD^{p/(1-p)})$ (for completeness, we give it in Appendix B).

Theorem 1.4. *Let $n, k \in \mathbb{R}^n$ where $k = \Omega(n)$ and let $p < 2$. There exists a deterministic, polynomial time algorithm which constructs a matrix $A \in \mathbb{R}^{n\times m}$ with the following properties: First, $m = \left\lceil n \cdot (k/n)^{\Omega(2-p)} \right\rceil$. Second, the number of non-zero entries in each row of A is $O\left((k/n)^2 \log(n/k)\right) = O(1)$. Finally, the Base Pursuit algorithm (w.r.t. A) has the ℓ_1/ℓ_p guarantee.*

1.3 Our techniques

We find a subspace in which the ratio between the ℓ_p and ℓ_r norms[4] of its vectors is the same up to some multiplicative constant, for any $0 < r < p < 2$. Specifically, for any vector x in the subspace $V \subseteq \mathbb{R}^n$ we shall have $||x||_r = \Omega(n^{1/r-1/p}||x||_p)$. We note that by Hölder's inequality, $||x||_r \leq n^{1/r-1/p}||x||_p$. For simplicity we focus on the case where $r = 1$. Dealing with general $r > 0$ requires little additional effort.

Before describing the construction method, we give a useful notation. A vector $x \in \mathbb{R}^n$ is said to be *p-spread* when no small set of its entries contain most of its ℓ_p mass. Another way to define the property of p-spreadness is the following: A vector x is not p-spread iff it can be written as $x = y + z$ where y is sparse and $||y||_p \gg ||z||_p$. It turns out that a vector x is p-spread iff $||x||_1 = \Omega(n^{1-1/p}||x||_p)$. In particular, if $||x||_1 \ll n^{1-1/p}||x||_p$ then $x = y + z$ where y is sparse and $||y||_p \gg ||z||_p$. The exact relation between the two notions is given in Section 3. We note that the notion of spreadness was first introduced by [GLR08] for $p = 2$ (along with the mentioned equivalence).

Our main result is a deterministic construction of an $\epsilon n \times n$ matrix A such that for any vector x for which $||x||_1 \ll n^{1-1/p}||x||_p$, it holds that $Ax \neq 0$. As a corollary we get that Ker(A) is a subspace of dimension $(1-\epsilon)n$ with constant distortion. To prove this property it suffices to prove two other properties. The first is that A does not expand the ℓ_p norm of any vector. Namely, $||Ax||_p = O(||x||_p)$ for any $x \in \mathbb{R}^n$. The second is that for any sufficiently sparse vector x, A is such that $||Ax||_p = \Omega(||x||_p)$. To see why these two properties derive the original one, assume that x is such that $||x||_1 \ll n^{1-1/p}||x||_p$. Then x is not p-spread, meaning that it can be split into $x = y + z$ where y is sparse and $||y||_p \gg ||z||_p$. Hence[5], $||A(y + z)||_p \geq ||Ay||_p - ||Az||_p > 0$ and x is not in the kernel of A.

The question remains how to construct such an "ℓ_p-norm preserving" matrix. The construction is done in several steps. First we construct a matrix A_0 of dimension $\epsilon n_0/2 \times n_0$ where $n_0 = \mathrm{poly}(\epsilon^{-1}) = O(1)$. A_0 will have similar but much stronger properties than those we would like to obtain for the final matrix A. We require A_0 to preserve the norm of any slightly non p-spread vector (as opposed to A which will preserve the norm of highly non p-spread vectors). Since

[3] Though it is not stated as a linear program, it can easily be transformed into one.

[4] We note that for $0 < r < 1$, the function $||\cdot||_r$ is not a norm. However, it is still well defined.

[5] We note that in the case where $p < 1$ the inequality does not hold. However, it does hold that $||A(y + z)||_p \geq p||Ay||_p - ||Az||_p > 0$ so the same arguments are still valid.

the size of A_0 is constant, we are able to use brute force methods in order to construct it (in $\exp(\mathrm{poly}(\epsilon^{-1})) = O(1)$ time). We note that this is the point in which our construction fails for $p = 2$ as the requirements of A_0 can only hold for[6] $p < 2$.

In the next step of the construction, we tensor A_0 with the identity matrix, thus obtaining a high dimension matrix $(A_1)'_{\epsilon n/2 \times n}$. Another way to think of the matrix A_1 is as the following linear operator. Consider a vector $x \in \mathbb{R}^n$. Let $I_1, \ldots, I_m$ be the partition of $[n]$ into $m = n/n_0$ equally sized consecutive intervals. Split x into m vectors of length n_0 corresponding to the intervals $I_1, \ldots, I_m$: $x_{I_1}, \ldots, x_{I_m}$. Then $A_1 x = (A_0 x_{I_1}, \ldots, A_0 x_{I_m})$. Notice that A_1 has the following property: For a vector $x \in \mathbb{R}^n$ that is not p-spread within all of the intervals, $||A_1 x||_p = \Omega(||x_p||)$. In fact, for the condition to hold it suffices that a constant percentage of the weight of x falls into non-spread intervals.

Clearly, not all sparse vectors have a constant fraction of their weight fall in non-spread intervals. For example, consider the vector $x = (1, \ldots, 1, 0, \ldots, 0)$ for which any x_I is either a zero vector or an all 1 vector. To overcome this difficulty and in order to deal with general sparse vectors we have an additional step in our construction. We concatenate two copies of the matrix, where the second copy has its rows permuted by some permutation, "uncorrelated" with the identity permutation. Namely, our final matrix A will be such that $Ax = (A_1 x, A_1 \pi x)$ where π is a permutation matrix. π is uncorrelated with the identity matrix in the following sense: For a sparse vector x, it cannot be the case that both x and πx are dense in many of the intervals $I_1, \ldots, I_m$. This permutation can be constructed from a bipartite constant degree unique neighbor expander. The details are given in Section 4.

This concludes the construction of the matrix but not the proof of its "norm preserving" property. The lack of correlation between the permutation π and the identity permutation ensures that most of the entries of x turn out to be in sparse intervals (i.e., in intervals having few non-zero entries). A sparse interval is also non-spread and thus its ℓ_p norm is preserved by the low dimension operator A_0. However, our difficulty that it might be the case that almost all of the weight of x is located exactly on entries that end up in dense intervals. To prove that such an event cannot occur we introduce the notion of *doomed indices*.

Consider the following partition of the (non-zero) entries of x: Assume w.l.o.g. that the lowest absolute value of a non-zero entry in x is 1. For each power of[7] 2 we define B_j as the set of indices in which $2^j \leq |x_i|^p < 2^{j+1}$. An entry $i \in B_j$ is *doomed* when the interval in which it lies contains too many entries from B_j. Namely, Let I_ℓ be the interval s.t. $i \in I_\ell$. Then i is *doomed* if $|I_\ell \cap B_j|$ is too large. The properties of the permutation π ensure that for either x or πx, a constant fracture of the weight lies in non-doomed entries. We will now show that a constant fracture of the weight of non-doomed entries (and thus of all entries) end up in non-spread intervals (either in x or πx). The conclusion would be that $||Ax||_p = ||A_1 x||_p + ||A_1 \pi x||_p = \Omega(||x||_p)$ since A_0 preserves the ℓ_p norm of non-spread vectors. Consider an interval I that is p-spread (that is, an interval for which x_I p-spread). Via a simple combinatorial argument, we prove that this implies that most of the weight of x_I originates from doomed indices.

1.4 Organization

In Section 2 we give some preliminaries, including some background regarding expander graphs. In Section 3 we formally explain the notion of well spread vectors and prove the equivalence between

[6] Specifically, for $p < 2$ we show that a random sign matrix A_0 has the required properties w.h.p. and that is not the case for $p = 2$.

[7] Actually we will eventually use a scaling factor of 4/3 and not 2.

a vector x being spread and holding $||x||_r = \Theta(n^{1/r-1/p}||x||_p)$. In Section 4 we describe the construction of the linear operator $A : \mathbb{R}^n \to \mathbb{R}^{\epsilon n}$. In Section 5 we prove the required properties of the mentioned permutation π. We prove a combinatorial weaker property in Section 5.1 and later the required properties in Section 5.2.

2 Preliminaries

For $n \in \mathbb{N}$ denote $[n] \overset{\Delta}{=} \{1, \ldots, n\}$. For a subset $S \subseteq [n]$ denote by $\bar{S}$ the complement of S, i.e., $\bar{S} \overset{\Delta}{=} [n] \setminus S$. In a graph G with vertex set V, for any subset $S \subseteq V$ of vertices we denote by $\Gamma_G(S)$ the set of S's neighbors. For $x \in \mathbb{R}^n$ and a subset $S = \{i_1 < i_2 < \ldots < i_{|S|}\} \subseteq [n]$ define x_S as the restriction of x to the indices of S. That is, $x_S \overset{\Delta}{=} (x_{i_1}, x_{i_2}, \ldots, x_{i_{|S|}})$. For $p > 0$ and $x \in \mathbb{R}^n$ we denote by $||x||_p$ the ℓ_p norm (or semi-norm when $p < 1$) of x. Namely, $||x||_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$. Denote by ℓ_p^n the n-dimensional vector space over $\mathbb{R}$, equipped with the ℓ_p-norm.

2.1 Expander Graphs

An undirected graph $G = (V, E)$ is called an (n, d, λ)-expander if $|V| = n$, the degree of each node is d and the second largest eigenvalue, in absolute value, of the adjacency matrix of G is λ. For any $d = p + 1$, where p is a prime congruent to 1 modulo 4, there are explicit constructions for infinitely many n of (n, d, λ)-expanders, where $\lambda \leq 2\sqrt{d-1}$ [Mar88, LPS88]. Expander graphs have an important property guaranteed by the commonly known expander mixing lemma. The expander mixing lemma states that, for any two subsets S, T of a regular expander graph G, the number of edges between S and T (denoted by $E(S, T)$) is approximately what you would expect in a random d-regular graph, i.e. $d|S||T|/n$. Formally put,

Lemma 2.1. *[Expander Mixing Lemma] Let $G = (V, E)$ be a d-regular graph with second-largest eigenvalue λ (in absolute value). Then for any two subsets $S, T \subseteq V$, let $E(S, T)$ denote the number of edges between S and T. We have*

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \leq \lambda\sqrt{|S||T|}$$

We require a bipartite expander with very good expansion properties that is regular on both sides. We obtain one by constructing the edge-vertex incidence graph of a spectral expander. Namely, given a spectral regular expander $G_0 = (V, F)$ with $|V| = m$ and degree d, we construct $G_1 = (L, R, E)$ in the following manner: $L = [m]$ corresponds to the set of vertices of G_0. $R = [nd/2]$ corresponds to the set of edges of G_0. Two vertices $v \in L$ and $e \in R$ are connected iff the edge e connects vertex v in G_0.

Lemma 2.2. *Let $G_0 = (V, F)$ be a regular spectral expander of degree d and second eigenvalue λ. Let $G_1 = (L, R, E)$ be the edge vertex adjacency graph of G_0. Let $T \subseteq L$. Denote by $\phi(T)$ the number of non-unique neighbors of T (i.e., the number of vertices in R with more than one neighbor in T). Then*

$$\phi(T) \leq d|T|\left(\frac{2|T|}{|L|} + \frac{2\lambda}{d}\right).$$

Proof. Denote by $\Gamma_{G_1}(T)$ the set of T's neighbors in R. The set T can also be viewed as a subset of V. Its neighbors in R (i.e., $\Gamma_{G_1}(T)$) correspond to the set of edges leaving the set T in G_0.

According to the expander mixing lemma (Lemma 2.1),

$$|\Gamma_{G_1}(T)| = d|T| - E_{G_0}(T,T) \overset{\text{E.M.L.}}{\geq} d|T|\left(1 - \frac{|T|}{m} - \frac{\lambda}{d}\right)$$

where $E_{G_0}(T,T)$ denotes the edges contained in the set T in G_0. The claim follows. □

3 Distortion and Well Spread Vectors

Let $0 < r < p < 2$. By Hölder's inequality, for any $x \in \mathbb{R}^n$ it holds that $||x||_r \leq n^{1/r-1/p}||x||_p$. Hence, to bound the distortion of a given subspace, it suffices to prove that for all of its vectors, $||x||_r = \Omega(n^{1/r-1/p}||x||_p)$. The following notion will ease us quantify this bound.

Definition 3.1. *Let $x \in \mathbb{R}^n$ and let $0 < r < p$.*

$$\Delta_{p\to r}(x) \triangleq \frac{n^{1/r-1/p}||x||_p}{||x||_r}$$

For briefness, for $p > 1$ we write $\Delta_p(x) \triangleq \Delta_{p\to 1}(x)$.

Notice that $\Delta_{p\to r}(x) \geq 1$. Hence, $\Delta_{p\to r}(V) \leq \max_{x\in V}\{\Delta_{p\to r}(x)\}$. In particular, in a subspace V where $\max_{x\in V}\{\Delta_{p\to r}(x)\} = O(1)$, the distortion is constant. We now give the formal definition and prove the equivalence of p-spreadness and low $\Delta_{p\to r}(\cdot)$ measure. We note that in [GLR08], the equivalent lemma was proved for $p = 2, r = 1$.

Definition 3.2. *Let $p \geq 1$. A vector $x \in \mathbb{R}^n$ is (p,α,η)-spread when for every $S \subseteq [n]$ with $|S| \leq \alpha n$, we have $||x_{\bar{S}}||_p \geq \eta||x||_p$ where $\bar{S} = [n] \setminus S$.*

Lemma 3.3. *Let $x \in \mathbb{R}^n$ and let $0 < r < p$.*

- *If x is (p,α,η)-spread then $\Delta_{p\to r}(x) \leq \eta^{-p/r}\alpha^{1/p-1/r}$.*
- *conversely, x is $(p, 2(\Delta_{p\to r}(x))^{-rp/(p-r)}, \Delta_{p\to r}(x)^{-1} \cdot (1-2^{1-r/p})^{1/r})$-spread.*

Proof. We first prove the first claim: Assume w.l.o.g. that $|x_1| \geq |x_2| \geq \ldots \geq |x_n|$ and that $||x||_p = 1$. For briefness, denote $y = x_{[\alpha n]}$, $z = x_{[\alpha n+1\ldots n]}$. According to our assumption, $||z||_p \geq \eta$. On the other hand, $||y||_p^p \leq ||x||_p^p = 1$, therefore $|x_{\alpha n}|^p \leq (\alpha n)^{-1}$ and thus $||z||_\infty \leq (\alpha n)^{-1/p}$.

$$||x||_r^r \geq ||z||_r^r \geq ||z||_p^p \cdot ||z||_\infty^{r-p} \geq \eta^p(\alpha n)^{(p-r)/p} = (\eta^p\alpha^{1-r/p}) \cdot n^{1-r/p}$$

This proves the first part of the claim.

We now prove the second claim: Let $||x||_p = 1$ and $\alpha = 2(\Delta_{p\to r}(x))^{-rp/(p-r)}$. Let $S \subseteq [n]$, $|S| \leq \alpha n$. By the definition of $\Delta_{p\to r}(\cdot)$, $||x||_r^r \geq n^{1-r/p}\Delta_{p\to r}(x)^{-r}$. On the other hand,

$$||x_S||_r^r \leq (\alpha n)^{1-r/p}||x_S||_p^r \leq (\alpha n)^{1-r/p} \leq n^{1-r/p}\Delta_{p\to r}(x)^{-r}\, 2^{1-r/p}$$

and

$$||x_{\bar{S}}||_r^r = ||x||_r^r - ||x_S||_r^r \geq \frac{n^{1-r/p}}{\Delta_{p\to r}(x)^r}(1-2^{1-r/p})$$

which implies $||x_{\bar{S}}||_p \geq ||x_{\bar{S}}||_r n^{1/p-1/r} \geq \Delta_{p\to r}(x)^{-1}(1-2^{1-r/p})^{1/r}$.

□

For our construction we require a matrix A_0, of constant dimensions, that preserves (up to some non-trivial distortion) the ℓ_p norm of any non p-spread vector. It turns out that a random (scaled) sign matrix has this property w.h.p.. As the dimensions of A_0 are constant, a random construction implies a deterministic one. The mentioned property can be verified by say, going over all the vectors in some ϵ-net of the ℓ_p unit sphere. As the dimensions are constant, the running time required for finding the matrix is also constant. We start with a formal definition of the required properties of the matrix.

Definition 3.4. *A matrix A of dimensions $\epsilon n \times n$ is called (p, α, β, τ)-norm preserving when it has the following properties:*

- *For any vector $x \in \mathbb{R}^n$, we have $||Ax||_p \leq \tau ||x||_p$*
- *For any vector $x \in \mathbb{R}^n$ s.t. x is not $(p, \alpha, 0.9)$-spread, $||Ax||_p \geq \beta ||x||_p$.*

The following Theorem states the existence of a norm preserving (scaled) sign matrix. Its proof is quite technical so we defer it to Appendix A. In a nutshell, we prove by standard methods, that a random sign matrix has some weaker norm preserving property for the ℓ_2 norm. We then show that this implies the required result for the ℓ_p norm.

Theorem 3.5. *Let $0 < \epsilon$, $p < 2$. There exists some $\alpha = \epsilon^{O((2-p)^{-1})}$ and $n_{initial} = \epsilon^{-O((2-p)^{-1})}$ with the following properties: For any $n \geq n_{initial}$, there exist a (scaled) sign matrix[8] $(A_0)_{\epsilon n \times n}$ s.t. A_0 is (p, α, β, τ)-norm preserving for some $\beta = n^{-O(1/p)}$ and $\tau = n^{O(1/p)}$.*

4 The Construction and Some Immediate Properties

In this section we present, for any $0 < r < p < 2$ and $\epsilon > 0$, a deterministic construction for a subspace $V \subseteq \ell_r^n$ with constant ℓ_p distortion and dimension $\dim(V) = (1-\epsilon)n$. Specifically, for any $\epsilon > 0$, we construct a matrix A of dimensions $\epsilon n \times n$ whose kernel is a subspace in which the ℓ_p distortion is $\epsilon^{-O_{p,r}(1)}$. We show that A preserves the ℓ_p norm of any sparse vector and prove this property ensures its kernel has the required property.

Let $0 < \epsilon < 1$. Recall that Theorem 3.5 guarantees the existence of a matrix $(A_0)_{0.5\epsilon n_0 \times n_0}$ that is (p, α, β, τ)-norm preserving, where $0 < \beta, \alpha < 1$, $\tau > 0$ and $n_0 \in \mathbb{N}$ are all functions of ϵ. This matrix can be found via exhaustive search in time, dependent only on ϵ. Let

$$A_1 = A_0 \bigotimes I_{2n/n_0 \times 2n/n_0}$$

where $\bigotimes$ is the tensor product and $n = n_0 m$ for some integer m. That is, for $b < 0.5\epsilon n_0$, $d < n_0$,

$$(A_1)_{a\lfloor 0.5\epsilon n_0 \rfloor + b, cn_0 + d} \triangleq \begin{cases} (A_0)_{b,d} & a = c \\ 0 & \text{otherwise} \end{cases}$$

We shall later define a pair of permutation matrices of dimension $n \times n$, denoted by π_1, π_2. The matrix we work with (whose kernel is the required subspace) is defined as follows:

$$A \triangleq A_1 \cdot \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix}.$$

Although we can assume w.l.o.g. that π_1 is the identity permutation, it is more convenient (for the purpose of the analysis) to discuss both permutations as arbitrary ones. As stated earlier, the permutations should be "uncorrelated" from one and other. We now give a formal definition to our requirements:

[8] If ϵn is not an integer, the number of rows is set to be $\lfloor \epsilon n \rfloor$

Definition 4.1. *Let $n_0, m \in \mathbb{N}$ and let $n = n_0 m$. Let $0 < \alpha, \gamma, \delta < 1$, $0 < p < 2$.*

- *Let $I_1 = [1, n_0], \ldots, I_m = [n - n_0 + 1, n]$.*
- *For each permutation π, we define an induced partition $(\mathcal{D}_\pi, \bar{\mathcal{D}_\pi})$ of the intervals according to some fixed $x \in \mathbb{R}^n$:*
$$\mathcal{D}_\pi \triangleq \{I \mid (\pi x)_I \text{ is } (p, \alpha, 0.9)\text{−spread}\}$$
- *Let $J_\pi = \bigcup_{I \notin \mathcal{D}_\pi} I$.*
- *a pair of permutations π_1, π_2 are called $(p, n_0, \alpha, \gamma, \delta)$-spreading if for any $x \in \mathbb{R}^n$ that is γ-sparse, it holds that*
$$||(\pi_1 x)_{J_{\pi_1}}||_p^p + ||(\pi_2 x)_{J_{\pi_2}}||_p^p \geq \delta ||x||_p^p.$$

The construction of the permutations is done via the following procedure: Let G_0 be an expander graph with m vertices of degree $2n_0$ and second eigenvalue $\lambda \leq 2\sqrt{2n_0 - 1} < 3\sqrt{n_0}$ (see Section 2.1)[9]. Let $G_1 = (L_1, R, E_1)$ be the edge-vertex adjacency graph of G_0 where $L_1 = [m]$ is the set of vertices and $R = [n]$ is the set of edges. Define $G_2 = (L, R, E)$ as the following graph: each vertex in L_1 is split into n_0 vertices. Every new vertex in L is connected to two edges (arbitrarily, say by lexicographic order) that were previously connected to the corresponding vertex in L_1. The graph $G_2 = (L, R, E)$ is a bipartite graph where in each side the degree of the vertices is 2. Let π_1' be a subset of edges that is a perfect matching between both sides (its existence is guaranteed by Hall's theorem and it can be obtained using e.g. the classic algorithm for finding a perfect matching in a bi-partite graph). Then $\pi_2' \triangleq E \setminus \pi_1'$ must also be a perfect matching. Since $|L| = |R| = n$, π_1', π_2' can be viewed as permutations over $[n]$. We denote these permutations as π_1, π_2.

Proving the spreading property of the permutations will be our main effort. We do so in Section 5. We now show that these properties indeed guarantee that the kernel of A has a low ℓ_p distortion. First, we show that the operator A indeed preserves the ℓ_p norm of any sparse vector.

Lemma 4.2. *Let $\epsilon > 0$ and let $A_0, A_1, A, \pi_1, \pi_2$ be as defined above (A_0 is (p, α, β, τ)-norm preserving). Let $\gamma > 0$. Assume that π_1, π_2 are $(p, n_0, \alpha, \gamma, \delta)$-spreading permutations.*

- *For any $x \in \mathbb{R}^n$ it holds that $||Ax||_p \leq 2^{1/p} \tau ||x||_p$.*
- *Assume that x is γ-sparse. Namely that it has at most γn non-zero entries. Then $||Ax||_p \geq \delta^{1/p} \beta ||x||_p$.*

Proof. The first part of the claim is trivial due to the definition of A and the norm preserving properties of A_0. To prove the second part of the claim, assume that x is γ-sparse. Since π_1, π_2 are spreading permutations,

$$||(\pi_1 x)_{J_{\pi_1}}||_p^p + ||(\pi_2 x)_{J_{\pi_2}}||_p^p \geq \delta ||x||_p^p$$

Hence, due to the properties of A_0,

$$||Ax||_p^p = ||A_1 \pi_1 x||_p^p + ||A_1 \pi_2 x||_p^p \geq \delta \cdot \beta^p ||x_S||_p^p.$$

□

[9] Notice that the mentioned expander graphs in Section 2.1 have particular restrictions regarding their degree ($2n_0$) and number of vertices (m). This does not undermine the correctness due to the following: First, the attributes we require of n_0 consist of a (constant) lower bound. Hence, the additional requirements of the expander construction can be held. Second, for any fixed n_0 and m where n_0 satisfies all the needed requirements, there exists some $m' = \Theta(m)$ for which the required expander can be constructed. Hence, if one wishes some specific value of n, we construct the matrix with some $n' \geq n$, $n' = O(n)$ (where the $O()$ is also independent of ϵ) and ignore the $n' - n$ last columns of the matrix. This leads to a matrix of size $O(\epsilon n) \times n$ and the analysis holds.

Theorem 4.3. *Let $\epsilon > 0$ and let $A_0, A_1, A, \pi_1, \pi_2$ be as defined above (A_0 is (p, α, β, τ)-norm preserving). Let $\gamma > 0$. Assume that π_1, π_2 are $(p, n_0, \alpha, \gamma, \delta)$-spreading permutations. Let $\tilde{p} = \min\{p, 1\}$ and let $x \in \mathbb{R}^n$ where*

$$\Delta_{p\to r}(x) > \gamma^{1/p-1/r} \cdot \tilde{p}^{-p/r}\delta^{-1/r}\beta^{-p/r}(\tilde{p}^p\delta\beta^p + 2\tau^p)^{1/r}.$$

Then $Ax \neq 0$.

Proof. $\Delta_{p\to r}(x)$ is sufficiently large so that x is not (p, γ, η)-spread for some $\eta < \tilde{p}\delta^{1/p}\beta(\tilde{p}^p\delta\beta^p + 2\tau^p)^{-1/p}$ (see Lemma 3.3). Assume w.l.o.g. that $||x||_p = 1$. Hence, there exist some subset $S \subseteq [n]$, $|S| \leq \gamma n$ where $||x_{\bar{S}}||_p^p < \eta^p$ and $||x_S||_p^p > 1 - \eta^p$. Since π_1, π_2 are spreading permutations, we have by Lemma 4.2 that

$$||Ax||_p^p \geq \tilde{p}^p||Ax_S||_p^p - ||Ax_{\bar{S}}||_p^p > \tilde{p}^p\delta\beta^p(1 - \eta^p) - 2\tau^p\eta^p = \tilde{p}^p\delta\beta^p - \eta^p(\tilde{p}^p\delta\beta^p + 2\tau^p) > 0$$

□

5 Analysis of the Permutations

5.1 Hashing Permutations

Before we prove the spreading properties of the permutations, we prove an easier combinatorial property of them which suffices for the case of vectors in $\{-1, 0, 1\}^n$.

Definition 5.1. *Let $n_0, m \in \mathbb{N}$ and let $n = n_0 m$. Let $0 < \mu, \gamma, \zeta < 1$.*

- *Let $I_1 = [1, n_0], \ldots, I_m = [n - n_0 + 1, n]$.*
- *For each permutation π, we define an induced partition $(\mathcal{E}_\pi, \bar{\mathcal{E}_\pi})$ of the intervals according to some $x \in \mathbb{R}^n$:*
$$\mathcal{E}_\pi \triangleq \{I \mid (\pi x)_I \text{ contains at least } \mu n_0 \text{ non zero}'s\}$$
- *Let $K_\pi = \bigcup_{I \notin \mathcal{E}_\pi} I$.*
- *a pair of permutations π_1, π_2 are called $(n_0, \mu, \gamma, \zeta)$-hashing if for any s-sparse vector $x \in \mathbb{R}^n$ with $s \leq \gamma n$, it holds that $(\pi_1 x)_{K_{\pi_1}}, (\pi_2 x)_{K_{\pi_2}}$ contain at least $2\zeta s$ non-zero entries (together). Namely, $||(\pi_1 x)_{K_{\pi_1}}||_0 + ||(\pi_2 x)_{K_{\pi_2}}||_0 \geq 2\zeta s$ (where $||y||_0$ denotes the number of non-zero entries in y).*

The following lemma essentially says that whenever $\gamma = O(\mu^2)$ and $n_0 = \Omega(\mu^{-2})$, the permutations are hashing.

Lemma 5.2. *Let $n_0 \in \mathbb{N}$ and $0 < \mu, \gamma, \zeta < 1$. Define $\xi \triangleq 16\mu^{-2}\gamma + 12\mu^{-1}n_0^{-1/2}$. If $2\zeta < 1 - \xi$ then π_1, π_2 are $(n_0, \mu, \gamma, \zeta)$-hashing permutations.*

Proof. Assume that π_1, π_2 are not hashing. We show that $1 - 2\zeta \leq \xi$. Let $x \in \mathbb{R}^n$ be a vector with respect to which the permutations are not hashing. Let $S \subset R$ (where $|R| = n$ is the right set of vertices of G_1) be a set of size $|S| \leq \gamma n$ containing the non zero entries of x. According to our (non hashing) assumption, $(\pi_1 x)_{K_{\pi_1}}, (\pi_2 x)_{K_{\pi_2}}$ contain less than $2\zeta|S|$ non zero's.

Let $T_1, T_2 \subseteq L_1$ be defined as the sets corresponding to $\mathcal{E}_{\pi_1}, \mathcal{E}_{\pi_2}$ (each interval is a left vertex in G_1). Define the following edge labelling in the graph G_1: Any edge is either in π_1 (viewed as a perfect matching) or π_2. Label the edge by the number 1 or 2 accordingly. Also, let $E_\ell(T,S)$ denote the set of edges connecting a set $T \subseteq L_1$ to a set $S \subseteq R$ labelled ℓ ($\ell \in \{1,2\}$). Notice that the vertices $i \in S$ that are connected to T_1 via an edge labelled 1 are exactly those not in K_{π_1} (i.e. in $S \setminus K_{\pi_1}$). An analog can be said for K_{π_2}. Hence,

$$|E_1(T_1,S)| + |E_2(T_2,S)| \geq (2-2\zeta)|S|.$$

Define $T = T_1 \cup T_2$. We get that the total number of edges from T to S is at least $(2-2\zeta)|S|$. As the right degree is 2, we get that (recall that $\phi(T)$ denotes the number of non-unique neighbors of T)

$$\phi(T) \geq (1-2\zeta)|S|.$$

We now express this as a bound involving $|T|$. Later we will use the expansion of the graph to upper bound $\phi(T)$ and get the required inequality. For each element of T_1, we have μn_0 elements of S that are "sent there" by the edges corresponding to π_1. Hence, $|T_1| \leq \frac{|S|}{\mu n_0}$. The same can be said for T_2 and thus $|T| \leq \frac{2|S|}{\mu n_0}$. We get that T has at least

$$(1-2\zeta)|S| \geq |T| \cdot ((1-2\zeta)\mu n_0/2) = 2n_0|T|(1-2\zeta)\mu/4$$

non-unique neighbors. On the other hand, by Lemma 2.2, T has at most

$$2n_0|T|\left(2n_0\frac{|T|}{n} + \frac{\lambda}{n_0}\right) \leq 2n_0|T|\left(4\mu^{-1}\gamma + 3n_0^{-1/2}\right)$$

non-unique neighbors (since $\lambda \leq 3\sqrt{n_0}$ and $|S| \leq \gamma n$). By combining inequalities, we get the required result.

$$1-2\zeta \leq 16\mu^{-2}\gamma + 12\mu^{-1}n_0^{-1/2} = \xi.$$

□

Corollary 5.3. *Let $0 < \mu < 1$. There exist some $n_{initial} = O(\mu^{-2})$ and $\gamma_{initial} = \Omega(\mu^{-2})$ with the following properties: For any $\gamma \leq \gamma_{initial}$ and $n_0 \geq n_{initial}$, we have that π_1, π_2 are $(n_0, \mu, \gamma, 1/3)$-hashing*

5.2 The Proof

We proceed to prove the $(p, n_0, \alpha, \gamma, \delta)$-spreading properties of π_1, π_2 for general vectors. The parameters $\alpha > 0$ and $0 < p < 2$ will be fixed and the other parameters will be set according to them. Let $x \in \mathbb{R}^n$ be a vector that is γ-sparse (the size restrictions of γ will be detailed later). Assume also w.l.o.g. that $||x||_p = 1$. Recall that we defined a partition of the indices into the intervals $I_1 = [1 \dots n_0], \dots I_m = [n - n_0 + 1 \dots n]$. We Define an additional partition of the indices (containing non-zero elements) into blocks. This partition will be dependent on the vector x. The granularity of the blocks shall depend on some $q > 1$ whose exact value will be given later.

Definition 5.4. *Let $i_0 \in [n]$ be an entry in which $|x_{i_0}|$ is non-zero and minimal w.r.t. all other absolute value of non-zero entries of x. Define for $j \geq 0$,*

$$B_j = \left\{ i \;\middle|\; |x_{i_0}|^p q^j \leq |x_i|^p < |x_{i_0}|^p q^{j+1} \right\}$$

Definition 5.5. *We say that a block B_j is ζ_ℓ-preserved by a permutation π_ℓ ($\ell = 1,2$) if $\pi_\ell(x_{B_j})$ has a ζ_ℓ percentage of its non-zero elements in α/r-sparse intervals $(I_1, \ldots, I_m)$, for*[10] $r \triangleq \left\lceil \log_q(\frac{(1-0.9^p)n_0}{0.9^p}) \right\rceil + \left\lceil \frac{1}{1-\frac{1}{q}} \right\rceil + 2$.

We require that $\gamma = O\left((\alpha/r)^2\right)$ and $n_0 = \Omega\left((\alpha/r)^{-2}\right)$ as in the requirements of Corollary 5.3 when applied for $\mu = \alpha/r$ (notice that although r depends on n_0, the dependence is logarithmic so the requirements holds for sufficiently large n_0). The following lemma is due to the hashing properties of the permutations:

Lemma 5.6. *Let B be some block that is ζ_ℓ-preserved by π_ℓ. It holds that*

$$\zeta_1 + \zeta_2 \geq 2/3.$$

Proof. Let y be the vector having the same entries as x in the indices of B and zero's elsewhere. Since x is γ-sparse, so is y. As γ is sufficiently small and n_0 is sufficiently large, by corollary 5.3 π_1, π_2 are $(n_0, \alpha/r, \gamma, 1/3)$-hashing. The claim easily follows. □

We now define the notion of doomed indices. Intuitively, these are indices that we are bound to "lose" as an interval containing them might become spread. We note that the definition is dependent on a specific vector $x \in \mathbb{R}^n$.

Definition 5.7. *Let $x \in \mathbb{R}^n$, let $i \in [n]$ be an index where $x_i \neq 0$ and let $\ell \in [2]$. Let B be the block containing i and let I be the interval s.t. $\pi_\ell(i) \in I$. The index i is called a doomed index w.r.t. π_ℓ (and x) if I contains more than $\alpha n_0/r$ indices from $\pi_\ell(B)$.*

The following lemma is an easy property of doomed indices

Lemma 5.8. *Let $D_\ell \subseteq [n]$ be the set of doomed indices w.r.t. π_ℓ ($\ell \in [2]$). Then*

$$\sum_{i \in D_1} |x_i|^p + \sum_{i \in D_2} |x_i|^p \leq 2 - \frac{2}{3q}.$$

Proof. According to Corollary 5.3, each block has 2/3 of its indices in sparse intervals (when counting each element 2 times, once in each permutation). In the worst case scenario, the indices that where "lost" in each block were of the largest possible value in comparison to those "saved". Namely, their ratio is q. This proves the claim. □

We now need to bound the weight of the non-doomed $|x_i|^p$'s in spread intervals. We do so by proving that in an interval I that is well spread, the relative weight of the doomed indices must be high. As the total weight of the doomed indices is bounded, we will get that the a constant fraction of the weight is present in non spread intervals.

Lemma 5.9. *Let $\pi = \pi_\ell$ be one of the permutations. Let I be an interval s.t. $(\pi x)_I$ is $(p, \alpha, 0.9)$-spread. Then*

$$||(\pi x)_{I \cap D_\ell}||_p \geq 0.9||(\pi x)_I||_p.$$

[10] The exact value of r comes from a calculation regarding the future equations. As q will be given a value independent of the other parameters, we get that $r = \Theta(\log(n_0))$.

Proof. Assume w.l.o.g. that $(\pi x)_I = (y_1, \dots, y_{n_0})$ where $|y_1| \geq |y_2| \geq \dots \geq |y_{n_0}|$. For briefness, define for each $i \in [n_0]$, $z_i = |y_i|^p$. We first mention that

$$z_{\alpha n_0} > \frac{0.9^p}{(1-0.9^p)n_0} \cdot z_1. \tag{1}$$

otherwise,

$$||(z_1, \dots, z_{\alpha n_0})||_1 \geq z_1 \geq z_{\alpha n_0} \frac{(1-0.9^p)n_0}{0.9^p} > \frac{(1-0.9^p)||(z_{\alpha n_0+1}, \dots, z_{n_0})||_1}{0.9^p}$$

and $||(z_{\alpha n_0+1}, \dots, z_{n_0})||_1 < 0.9^p ||(z_1, \dots, z_{n_0})||_1$, meaning that the interval is not $(p, \alpha, 0.9)$ spread. We consider a block B to be doomed w.r.t. an interval I and a permutation π when at least $\alpha n_0 / r$ of its indices are sent by π to I. Inequality 1 indicates that there are at most

$$\log_q \left(\frac{(1-0.9^p)n_0}{0.9^p} \right) + 1 \leq r - \left\lceil \frac{1}{1-\frac{1}{q}} \right\rceil - 1$$

many non-doomed blocks w.r.t. π, I having their elements appear in $\{z_1 \dots z_{\alpha n_0}\}$. As $z_1, \dots, z_{\alpha n}$ are all non-zero, it follows that at least

$$\alpha n_0 - \left(r - \left\lceil \frac{1}{1-\frac{1}{q}} \right\rceil - 1 \right) \frac{\alpha n_0}{r} = \frac{\alpha n_0}{r} \left(\left\lceil \frac{1}{1-\frac{1}{q}} \right\rceil + 1 \right)$$

elements of $\{z_1, \dots, z_{\alpha n}\}$ originate from a doomed index. Denote by d_{big} (g_{big}) the sum of z_i's originating from doomed (non-doomed) indices in $\{z_1 \dots z_{\alpha n_0}\}$ and by d_{small} (g_{small}) the sum of z_i's originating from doomed (non-doomed) indices in $\{z_{\alpha n_0+1} \dots z_{n_0}\}$. We get that

$$d_{big} \geq \frac{z_{\alpha n_0} \alpha n_0}{r} \left(\left\lceil \frac{1}{1-\frac{1}{q}} \right\rceil + 1 \right).$$

Also,

$$g_{small} \leq z_{\alpha n_0} \cdot \frac{\alpha n_0}{r} \cdot (2 + q^{-1} + q^{-2} + \dots) \leq d_{big}.$$

This holds since: In the worst case scenario (i.e., g_{small} gets the maximal value), the maximum valued element in $\{z_{\alpha n_0+1} \dots z_{n_0}\}$ has the value of $z_{\alpha n_0}$ and appears $2\alpha n_0 / r$ times (as the smallest and largest element in two consecutive non-doomed blocks). After this, each element must be q times smaller than the previous and may appear at most $\frac{\alpha n_0}{r}$ times. Concluding, we get that

$$\frac{||(\pi x)_{I \cap D_\ell}||_p^p}{||(\pi x)_I||_p^p} = \frac{d_{small} + d_{big}}{||z||_1} \geq \frac{d_{small} + g_{small}}{||z||_1} =$$

$$\frac{||(z_{\alpha n_0+1}, \dots, z_{n_0})||_1}{||z||_1} \geq 0.9^p.$$

□

By setting $q = 4/3$ we get the following corollary

Corollary 5.10. *For $\ell = 1, 2$, let (recall Definition 4.1) $\mathcal{D}_{\pi_\ell}$ be the set of intervals s.t. $(\pi_\ell x)_I$ is $(p, \alpha, 0.9)$-spread. Then*

$$\sum_{I \in \mathcal{D}_{\pi_1}} ||(\pi x)_I||_p^p + \sum_{I \in \mathcal{D}_{\pi_2}} ||(\pi x)_I||_p^p \leq 1.86 ||x||_p^p$$

Proof. Assume w.l.o.g. that $||x||_p = 1$. By Lemma 5.8, the sum of doomed $|x_i|^p$'s (counting each i twice, once for each π_ℓ) is at most 1.5. By Lemma 5.9 and $p < 2$ we have that,

$$\sum_{I \in \mathcal{D}_{\pi_1}} ||(\pi x)_I||_p^p + \sum_{I \in \mathcal{D}_{\pi_2}} ||(\pi x)_I||_p^p \leq \frac{1.5}{0.9^p} < 1.86$$

□

Corollary 5.11. *Let* $0 < \alpha < 1$. *Let* $n \in \mathbb{N}$ *be some sufficiently large integer. There exist some* $n_{initial} = O(\alpha^{-2} \log^2(\alpha^{-1}))$ *and* $\gamma = \Omega(\alpha^2 \log^{-2}(\alpha^{-1}))$ *with the following properties: Let* $n_0 \geq n_{initial}$, *and let* π_1, π_2 *be a pair of permutations constructed as in Section 4 w.r.t mentioned parameters. Then* π_1, π_2 *are* $(p, n_0, \alpha, \gamma, 0.14)$*-spreading.*

Theorem 1.2 easily follows from the above Corollary, Theorem 3.5 and Theorem 4.3.

6 Acknowledgements

I would like to thank Amir Shpilka, Yuval Rabani and Gideon Shechtman for their helpful comments and and stimulating conversations.

References

[AAM05] S. Artstein-Avidan and V.D. Milman. Logarithmic reduction of the level of randomness in some probabilistic geometric constructions. *Journal of Functional Analysis*, 236:297–329, 2005. 1, 1

[Bar07] R.G. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 24(4):118, 2007. 1.2.1

[CT06] E. J. Candés and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006. 1.2.1

[DeV07] R. DeVore. Deterministic constructions of compressed sensing matrices. *J. of computational complexity*, 23:918–925, 2007. 1

[DS89] D.L. Donoho and P.B. Stark. Uncertainty principles and signal recovery. *SIAM Journal on Applied Mathematics*, 49(3):906–931, 1989. 1

[Dvo59] A. Dvoretzky. A theorem on convex bodies and applications to banach spaces. *Proceedings of the National Academy of Sciences of the United States of America*, 45(2):223–226, 1959. 1

[FLM77] T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Mathematica*, 139(1):53–94, 1977. 1

[GI10] A. Gilbert and P. Indyk. Sparse recovery using sparse matrices. In *Proceedings of the IEEE*, volume 98, pages 937 – 947, 2010. 1.2.1

[GLR08] V. Guruswami, J. R. Lee, and A. A. Razborov. Almost Euclidean subspaces of l_1^n via expander codes. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 353–362, 2008. 1, 1, 1.1, 1.3, 3

[GLW08] V. Guruswami, J. R. Lee, and A. Wigderson. Euclidean sections of with sublinear randomness and error-correction over the reals. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques, 11th International Workshop, APPROX 2008, and 12th International Workshop, RANDOM 2008*, pages 444–454, 2008. 1, 1, 1.2.1

[Ind06] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006. 1, 1

[Ind07] P. Indyk. Uncertainty principles, extractors, and explicit embeddings of ℓ_2 into ℓ_1. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pages 615–620, 2007. 1, 1

[IS10] P. Indyk and SJ Szarek. A simple construction of almost-Euclidean subspaces of via tensor products. *e-print arXiv*, 1001, 2010. 1, 1

[Kas77] B. S. Kashin. Diameters of some finite-dimensional sets and classes of smooth functions. *Izv. AN SSSR. Ser. Mat.*, 41(2):334351, 1977. 1

[KT07] B.S. Kashin and V.N. Temlyakov. A remark on compressed sensing. *Mathematical notes*, 82(5):748–755, 2007. 1.2.2

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. 1

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988. 2.1

[LS08] S. Lovett and S. Sodin. Almost Euclidean sections of the N-dimensional cross-polytope using $O(N)$ random bits. *Commun. Contemp. Math.*, 10(4):477–489, 2008. 1, 1, A

[Mar88] G. A. Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problems of Information Transmission*, 24(1):39–46, 1988. 2.1

[Rud60] W. Rudin. Trigonometric series with gaps. *J. Math. Mech*, 9(2):203–227, 1960. 1

[Sch84] C. Schutt. Entropy numbers of diagonal operators between symmetric Banach spaces. *Journal of approximation theory*, 40(2):121–128, 1984. A

A A Low Dimension Norm Preserving Matrix

We start by presenting a Lemma showing that a random (scaled) sign matrix preserves the ℓ_2 norm of any vector x with large $\Delta_{2\to 1}(\cdot)$ measure. We shall then see that a vector x that is not p-spread (as in the requirements of Theorem 3.5) has this property (for $p < 2$).

The following result is standard, given covering estimates of Schütt [Sch84]. In a nutshell, one proves that a random (scaled) sign matrix of dimension $\epsilon n \times n$ preserves the norm of any fixed vector of $\mathbb{R}^n$ with probability $1 - \exp(\epsilon n)$. One then proceeds to union bound over a net, covering all vectors with large $\Delta_{2\to 1}(\cdot)$ measure. For a complete proof see e.g. [LS08], Lemma B.

Lemma A.1. *Let $0 < \epsilon < 1$. There exist some $D = O(\epsilon^{-1/2}\log(\epsilon^{-1}))$ and $n_{initial} = \epsilon^{-O(1)}$ such that for any $n \geq n_{initial}$ there exists a (scaled) sign matrix $(A_0)_{\epsilon n \times n}$ with the following properties:*

1. *For any vector $x \in \mathbb{R}^n$ s.t. $\Delta_{2\to 1}(x) > D$, $||A_0 x||_2 \geq \beta' ||x||_2$ for some $\beta' = \Omega(1)$.*
2. *For any vector $x \in \mathbb{R}^n$, $||A_0 x||_2 \leq \tau' ||x||_2$ for some $\tau' = O(\epsilon^{-1/2})$.*

We now prove a connection between ℓ_p spreadness and ℓ_2 spreadness. As a corollary we get a connection between the ℓ_p-spreadness and the $\Delta_{2\to 1}(\cdot)$ measure.

Lemma A.2. *Let $0 < p < 2$. Let $x \in \mathbb{R}^n$ be some vector that is not $(p, \alpha, 0.9)$-spread. Then x is not $(2, \sqrt{\alpha}, 10^{1/p}\alpha^{1/(2p)-1/4})$-spread.*

Proof. Assume w.l.o.g. that $|x_1| \geq |x_2| \geq \ldots \geq |x_n|$ and that $||x||_p^p/n = \mathbb{E}_{i\in[n]}[|x_i|^p] = 1$. Let $0 < z < 1$ be a parameter. First, notice that for any $z > 0$, $|x_{\alpha^z n}|^p \leq \alpha^{-z}$. That is, $|x_{\alpha^z n}| \leq \alpha^{-z/p}$. Now,

$$||x_{[\alpha^z n+1\ldots n]}||_2^2 \leq ||x_{[\alpha^z n+1\ldots n]}||_\infty^{2-p} \cdot ||x_{[\alpha^z n+1\ldots n]}||_p^p \leq$$
$$\alpha^{-(2-p)z/p} \cdot ||x_{[\alpha n+1\ldots n]}||_p^p \leq \alpha^{z(1-2/p)} \cdot 0.9n.$$

On the other hand,

$$||x_{[\alpha^z n]}||_2^2 \geq ||x_{[\alpha n]}||_2^2 = \alpha n \cdot \mathbb{E}_{i\in[\alpha n]}[x_i^2] \geq \alpha n \cdot (\mathbb{E}_{i\in[\alpha n]}[|x_i|^p])^{2/p} \geq$$

$$\alpha n \left(\frac{0.1n}{\alpha n}\right)^{2/p} \geq 0.01^{1/p}\alpha^{1-2/p} n.$$

By combining both equations we get

$$\frac{||x_{[\alpha^z n+1 \ldots n]}||_2^2}{||x||_2^2} \leq \frac{||x_{[\alpha^z n+1 \ldots n]}||_2^2}{||x_{[\alpha^z n]}||_2^2} \leq \frac{0.9\alpha^{z(1-2/p)}}{0.01^{1/p}\alpha^{1-2/p}} \leq 100^{1/p}\alpha^{(2/p-1)(1-z)}$$

and by setting $z = 1/2$ we get the required result.

□

The following is an immediate corollary of the above Lemma and Lemma 3.3.

Corollary A.3. *Let $0 < p < 2$, $\alpha > 0$, and let $x \in \mathbb{R}^n$ be some vector that is not $(p, \alpha, 0.9)$-spread. Then $\Delta_{2\to 1}(x) \geq \alpha^{1/4-1/(2p)}/(4 \cdot 10^{1/p})$.*

Hence, by setting $\alpha = \epsilon^{O((2-p)^{-1})}$ (so that $\alpha^{1/4-1/(2p)}/(4 \cdot 10^{1/p}) = D$ where D is as in Lemma A.1), one can achieve a matrix A_0 with the following properties: Let n, β', τ' be as in Lemma A.1 and let $x \in \mathbb{R}^n$:

- $||A_0 x||_2 \leq \tau' ||x||_2$.
- Assume $\Delta_{2\to 1}(x) \geq \alpha^{1/4-1/(2p)}/(4 \cdot 10^{1/p})$. Then $||A_0 x||_2 \geq \beta' ||x||_2$

We now prove that this matrix has the required properties of Theorem 3.5. Let $x \in \mathbb{R}^n$. Then

$$||A_0 x||_p \leq (\epsilon n)^{1/p-1/2}||A_0 x||_2 \leq (\epsilon n)^{1/p-1/2}\tau'||x||_2 \leq (\epsilon n)^{1/p-1/2}\tau'||x||_p \triangleq \tau||x||_p.$$

Assume now that x is not $(p, \alpha, 0.9)$-spread. Then by Corollary A.3, $\Delta_{2\to 1}(x) \geq \alpha^{1/4-1/(2p)}/(4 \cdot 10^{1/p})$ and

$$||A_0 x||_p \geq ||A_0 x||_2 \geq \beta'||x||_2 \geq \beta' n^{1/2-1/p}||x||_p \triangleq \beta \cdot ||x||_p$$

Hence, A_0 is (p, α, β, τ)-norm preserving for $\beta = n^{-O(1/p)}$, $\tau = n^{O(1/p)}$ (as $n \geq \epsilon^{-1}$).

B Proof of the ℓ_1/ℓ_p guarantee

Fix some $1 < p < 2$. Let $V \subseteq \mathbb{R}^n$ be a subspace with $\Delta_{1\to p}(V) \leq D$. Namely, for any $x \in V$ it holds that $||x||_p \leq n^{1-1/p}D||x||_1$.

Lemma B.1. *Let $0 \neq x \in V$. Then $||x||_0 \geq n/D^{p/(p-1)}$.*

Proof. Let $S \subseteq [n]$ be the set of non-zero indices of x. We have,

$$||x||_1 = \sum_{i\in S}|x_i| \leq |S|^{1-1/p}\left(\sum_{i\in S}|x_i|^p\right)^{1/p} = |S|^{1-1/p}||x||_p \leq |S|^{1-1/p}\left(n^{1/p-1}D||x||_1\right)$$

where the last equality holds since $x \in V$. As $||x||_1 > 0$ we get that $|S| \geq n \cdot D^{p/(1-p)}$. □

Lemma B.2. *Let $0 \neq x \in V$. Then for any $S \subseteq [n]$ where $|S| < n(2D)^{p/(1-p)}$, $||x_S||_1 < ||x||_1/2$.*

Proof. Similarly to the previous lemma,

$$||x_S||_1 = \sum_{i \in S} |x_i| \leq |S|^{1-1/p} \left(\sum_{i \in S} |x_i|^p \right)^{1/p} = |S|^{1-1/p} ||x_S||_p \leq$$

$$|S|^{1-1/p} ||x||_p \leq |S|^{1-1/p} \left(n^{1/p-1} D ||x||_1 \right) < ||x||_1/2$$

□

Lemma B.3. *Suppose $u \in \mathbb{R}^n$, $||u||_0 < n(2D)^{p/(1-p)}$. Then for any $v = u + x$, where $x \in V$, $x \neq 0$,*

$$||v||_1 > ||u||_1$$

Proof. Let S be the set of indices in which u is non-zero. Then

$$||v||_1 = \sum_{i \in [n]} |u_i + x_i| \geq \sum_{i \in S} |u_i| - |x_i| + \sum_{i \notin S} |x_i| = ||u||_1 + ||x||_1 - 2||x_S||_1 > ||u||_1$$

□

Recall that the “Base Pursuit” algorithm finds the vector u given Au (where $\mathrm{Ker}(A) = V$) via the following linear program:

$$\min \ ||y||_1 \quad \text{s.t. } Ay = Au.$$

The previous lemma guarantees that the output is exactly u provided $||u||_0 < n(2D)^{p/(1-p)}$. We now prove the robustness (i.e., ℓ_1/ℓ_p guarantee) of the algorithm. Let $u \in \mathbb{R}^n$ be some vector and let y be the solution of the above linear program. Notice that

$$y = u + \mathrm{argmin}_{x \in V} ||u + x||_1.$$

Theorem B.4. *Let $u \in \mathbb{R}^n$ and let u' be such that $||u'||_1 \leq ||u||_1$ and $u - u' \in V$. Let $k = \lfloor (4D)^{p/(1-p)} n \rfloor$. Then,*

$$||u - u'||_1 \leq 4\sigma_k(u)_1$$

and

$$||u - u'||_p \leq k^{1/p-1} \sigma_k(u)_1$$

Proof. Since $u - u' \in V$ the second inequality stems from the first. Let $S \subseteq [n]$ be of size $|S| = k$ containing the largest entries of u in absolute value. Let $\bar{S} = [n] \setminus S$. First,

$$\sigma_k(u)_1 = ||u - u_S||_1 = ||u_{\bar{S}}||_1.$$

Now,

$$||u - u'||_1 = ||(u - u')_S||_1 + ||(u - u')_{\bar{S}}||_1 \leq ||(u - u')_S||_1 + ||u_{\bar{S}}||_1 + ||(u')_{\bar{S}}||_1.$$

As $||u'||_1 \leq ||u||_1$,

$$||(u')_{\bar{S}}||_1 - ||u_{\bar{S}}||_1 = ||u'||_1 - ||u||_1 - ||(u')_S||_1 + ||u_S||_1 \leq ||(u - u')_S||_1.$$

Hence,

$$||(u')_{\bar{S}}||_1 \leq ||u_{\bar{S}}||_1 + ||(u - u')_S||_1$$

and

$$||(u - u')||_1 \leq 2||(u - u')_S||_1 + 2||u_{\bar{S}}||_1$$

As $u - u' \in V$ we get

$$||(u-u')_S||_1 \leq |S|^{1-1/p}||(u-u')_S||_p \leq |S|^{1-1/p}||u-u'||_p \leq |S|^{1-1/p}n^{1/p-1}D||u-u'||_1.$$

Our assumption on $|S|$ guarantees that $|S|^{1-1/p}n^{1/p-1}D \leq 1/4$. It follows that

$$||u-u'||_1 \leq ||u-u'||_1/2 + 2||u_{\bar{S}}||_1$$

which gives

$$||u-u'||_1 \leq 4||u_{\bar{S}}||_1 = 4\sigma_k(u)_1$$

□

Corollary B.5. *Let $u \in \mathbb{R}^n$ and let y be the output of the "Base Pursuit" algorithm. Let $k = \lfloor (4D)^{p/(1-p)}n \rfloor$. Then*

$$||u-y||_1 \leq 4\sigma_k(u)_1$$

and

$$||u-y||_p \leq k^{1/p-1}\sigma_k(u)_1$$

ECCC ISSN 1433-8092
http://eccc.hpi-web.de