Electronic Colloquium on Computational Complexity, Report No. 12 (2007)

Almost Euclidean sections of the N-dimensional cross-polytope using O(N) random bits

Shachar Lovett[1] and Sasha Sodin[2]

January 23, 2007

Abstract

It is well known that $\mathbb{R}^N$ has subspaces of dimension proportional to N on which the ℓ_1 norm is equivalent to the ℓ_2 norm; however, no explicit constructions are known. Extending earlier work by Artstein–Avidan and Milman, we prove that such a subspace can be generated using $O(N)$ random bits.

1 Introduction

We study embeddings of ℓ_2 spaces into ℓ_1 spaces. Recall that the ℓ_p norm on $\mathbb{R}^N$ is defined by:

$$\|x\|_p = \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p} \quad (p \geq 1)$$

The following inequality holds on $\mathbb{R}^N$:

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{N} \|x\|_2$$

It is well known since the work of Figiel, Lindenstrauss and Milman [7] and Kashin [13] that there exists a subspace E of $\mathbb{R}^N$ of dimension

[1]Faculty of Mathematics and Computer Science, The Weizmann Institute of Science, POB 26, Rehovot 76100, Israel. Email: Shachar.Lovett@weizmann.ac.il

[2]School of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, 69978, Israel. Email: sodinale@tau.ac.il

ISSN 1433-8092

$\Theta(N)$ such that for all $x \in E$, $\|x\|_1 = \Theta(\sqrt{N}\|x\|_2)$ (for the convenience of the reader, we recall the Θ-notation at the end of the introduction).

More formally put, for every $0 < \eta < 1$ and every $N \in \mathbb{N}$ (large enough), there exists an ηN-dimensional subspace $E \hookrightarrow \mathbb{R}^N$ such that for every $x \in E$:

$$c_\eta \sqrt{N}\|x\|_2 \leq \|x\|_1 \leq \sqrt{N}\|x\|_2 \tag{1}$$

where $c_\eta > 0$ depends only on η.

The subspace E gives in particular an embedding of $(\mathbb{R}^{\eta N}, \|\cdot\|_2)$ into $(\mathbb{R}^N, \|\cdot\|_1)$. This allows to reduce various problems in ℓ_2 norm to corresponding problem in ℓ_1 norm, with only a constant blowup in the dimension.

An explicit construction of E would therefore have various algorithmic applications. This was put forward by Indyk [10, 11], who proved several related results and applied them to problems in Computer Science.

No explicit subspace E satisfying (1) has been found so far (for large N). However, it is known that a randomly chosen subspace, under various natural definitions of distributions of subspaces, satisfies (1) with probability very close to 1.

In a sense, this situation is typical for various problems in asymptotic convex geometry, as for numerous properties satisfied by "random" high-dimensional objects it is hard to generate a deterministic object satisfying the property.

To resolve this dissonance, a new line of research was introduced by Sh. Artstein-Avidan and V. Milman. In the innovating work [3], the authors proposed to reduce the randomness needed to generate the random objects. More precisely, they showed that the random constructions in the proofs of a broad range of theorems, from Milman's Quotient of Subspace theorem to Zig-Zag approximation, can be performed on the finite probability space $\{-1,+1\}^R$ equipped with the uniform probability measure, where $R \in \mathbb{N}$ is reasonably small (the reader may refer to the work [4] by Artstein–Avidan and Milman for further developments and to the ICM lecture by Szarek [16] for a discussion of these and related issues).

In this case, we say informally that R random bits are used in the construction. For example, regarding the property (1), Artstein-Avidan and Milman showed that $O(N \log N)$ random bits suffice to construct the subspace E.

Their proof uses ε-net arguments, and decreasing the number of random bits beyond $\Omega(N)$ will probably require entirely new proof ideas. However, the $\log N$ factor in [3] seemed to be an artefact of the proof.

In this work, we show that this is indeed the case, and reduce the number of random bits to $O(N)$ using a modification of the construction from [3].

Theorem 1. *For every $0 < \eta < 1$, an ηN-dimensional subspace of $\mathbb{R}^N$ satisfying (1) can be generated using $O(N)$ random bits. Moreover, the memory needed to generate the subspace is $O(\log^2 N)$.*

As promised, we recall now the Θ-notation:

Notation. Let f, g be two functions from $(a, +\infty)$ or $(a, +\infty) \cap \mathbb{N}$ to $\mathbb{R}_+$. We will write:

1. $f = O(g)$ if there exist two constants $C > 0$ and $x_0 \geq a$ such that $f(x) \leq Cg(x)$ for every $x \geq x_0$;
2. $f = o(g)$ if $f(x)/g(x) \to 0$ as $x \to \infty$;
3. $f = \Omega(g)$ if $g = O(f)$;
4. $f = \omega(g)$ if $g = o(f)$;
5. and finally, $f = \Theta(g)$ if $f = O(g)$ and $f = \Omega(g)$.

Acknowledgement. We thank our supervisors, Omer Reingold and Vitali Milman, for constant support and for their interest in this work. We are also grateful to Shiri Artstein–Avidan for numerous discussions and explanations, and in particular for focusing our attention on bounding the operator norm as the main technical challenge.

2 Construction

Denote $\xi = 1 - \eta$, $n = \xi N$. We will construct a random $n \times N$ sign matrix A (that is, $A_{ij} = \pm 1$) using $O(N)$ random bits, and then prove that the kernel

$$E = \mathrm{Ker} A = \left\{ x \in \mathbb{R}^N \,\middle|\, Ax = 0 \right\}$$

satisfies (1) with high probability.

Recall the following simple definition:

Definition 1. The Hadamard (or entrywise) product of two $n \times N$ matrices A_1 and A_2 is the $n \times N$ matrix $A = A_1 \bullet A_2$, defined by $(A)_{i,j} = (A_1)_{i,j}(A_2)_{i,j}$.

Our random matrix A will be the Hadamard product $A_1 \bullet A_2$ of two random matrices A_1 and A_2, independent of each other. The construction of A_1 and A_2 will use two different techniques, both of them quite common.

Definition 2. A sequence of random variables $X_1, ..., X_M$ is called k-wise independent if every k of them are independent.

It is well-known that it is possible to construct M k-wise independent random signs from $O(k \log M)$ truly independent random signs. More formally, we have:

Lemma A. For every $k \leq M$, there exists a subset

$$\Upsilon_{k,M} \subset \{-1, 1\}^M$$

such that $|\Upsilon_{k,M}| = 2^{C_{k,M}}$, $C_{k,M} = O(k \log M)$, and for the randomly chosen vector $X = (X_1, ..., X_M)$ from $\Upsilon_{k,M}$, the following properties hold:

1. For $1 \leq m \leq M$, $\mathbb{P}\{X_m = -1\} = \mathbb{P}\{X_m = 1\} = 1/2$.
2. The coordinates of X are k-wise independent.
3. The set $\Upsilon_{k,M}$ is *explicit*, meaning that there exists a bijection $\upsilon_{k,M} : \{-1, 1\}^{C_{k,M}} \to \Upsilon_{k,M}$ that can be computed in time polynomial in k and M.

Definition 3. The random variables $(X_1, \cdots, X_M)$ satisfying the conditions 1.-2. of Lemma A are called k-wise independent random signs.

For completeness, we reproduce a proof of Lemma A due to Alon, Babai and Itai [1] in Appendix A.

The elements of our first matrix A_1 will be k-wise independent with $k = \Theta(\log N)$. That is, A_1, regarded as a vector in $\{-1, 1\}^{nN}$, will be a uniformly chosen element of $\Upsilon_{k,nN}$.

Remark. Regardless of the distribution of the random sign matrix A_2, the entries A_{ij} of the Hadamard product $A = A_1 \bullet A_2$ are k-wise independent random signs (in the sense of Definition 3).

Recall the definition of ℓ_2 operator norm:

Definition 4. For a matrix A, we define its operator norm as

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \ .$$

The k-wise independence of the elements of A_1 allows to control the operator norm of A. The following technical lemma may be of independent interest:

Lemma 2. *Let V be any $n \times N$ matrix of $2k$-wise independent random signs, $k \leq c_2\sqrt{N}$ (where $c_2 > 0$ is a numerical constant). Denote $\xi = n/N \leq 1$. Then, for $t \geq 0$,*

$$\begin{aligned}\mathbb{P}\left\{\frac{1}{\sqrt{N}}\|V\| \geq 1 + \sqrt{\xi} + t\right\} &\leq 2n\left(1 + \frac{t}{1+\sqrt{\xi}}\right)^{-2k} \\ &\leq 2n \exp\left\{\frac{-2kt}{1+\sqrt{\xi}+t}\right\} \ .\end{aligned}$$

We prove the lemma in Section 3.

Corollary 3. *Let $0 < \xi < 1$, $n = \xi N$; let A_1 be constructed as above with k-wise independent entries, and let $A = A_1 \bullet A_2$, where A_2 is an arbitrary random sign matrix independent of A_1. There exists a numerical constant $C_1 > 0$ such that for $k \geq C_1 \log n$,*

$$\mathbb{P}[\|A\| > 3\sqrt{N}] < 1/n \ .$$

We now head to construct a probability space for A_2; we use random walks on expander graphs (see Hoory, Linial and Wigderson [9] for an extensive survey). Let us recall the basic definitions.

Let $G = (\mathcal{V}, \mathcal{E})$ be a d-regular graph; the value of d plays no significant role in the estimates, so the reader may assume $d = 4$. Let P^G be the transition matrix of the random walk of G:

$$P^G_{uv} = \begin{cases} 1/d, & (u,v) \in \mathcal{E} \\ 0, & (u,v) \notin \mathcal{E}. \end{cases}$$

Denote by $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots$ the eigenvalues of P^G arranged in decreasing order, and denote $\lambda = \max_{i \geq 2} |\lambda_i|$.

In this notation, the graph G is called a $(|\mathcal{V}|, d, \lambda)$-graph. We will only need the following fact (cf. [9], [3]):

Fact. For any $d \geq 3$ and any number of vertices $|\mathcal{V}|$ (big enough), there exists a $(|\mathcal{V}|, d, \lambda)$-graph $G = (\mathcal{V} = \{1, 2, \cdots, |\mathcal{V}|\}, \mathcal{E})$ such that

1. $\lambda < 0.95$ and
2. G is *explicit*, formally meaning that set of neighbours

$$\{u \in \mathcal{V} \,|\, (u,v) \in \mathcal{E}\}$$

of any vertex $v \in \mathcal{V}$ can be computed in time that is polynomial in $\log |\mathcal{V}|$.

Sometimes we will call such a graph an expander graph with parameter λ.

Let $G = (\mathcal{V}, \mathcal{E})$ be an expander graph, with vertices $\mathcal{V}$ indexed by the elements of $\Upsilon_{4,N}$. Let $v_1, v_2, \cdots, v_n$ be a random walk of length n in G, starting from a random element of $\mathcal{V}$. Write the sign vectors corresponding to $v_1, \cdots, v_n$ in $\Upsilon_{4,N}$ as the rows of A_2.

The use of expander graphs is similar to [3]; however, we use constant degree expanders. We also show it suffices to use 4-wise independent rows rather than truly independent rows. This enables the computation to be performed using less memory ($O(\log^2 N)$).

Note that the construction uses in total

$$\begin{aligned} O(\log n \log(Nn)) + O(\log N) + O(n \log d) \\ = O(n + \log n \log N) = O(N) \end{aligned} \tag{2}$$

random bits. Also, we have the following:

Lemma 4. *Let A_1 be any constant sign matrix, and let A_2 be constructed as above. For every $x \in \mathbb{R}^N$ and any $\varepsilon \leq c_\lambda \sqrt{\xi}$,*

$$\mathbb{P}\left\{ \|Ax\|_2 < 6\varepsilon\sqrt{N}\|x\|_2 \right\} < C_\lambda p_\lambda^n \ ,$$

where the constants $C_\lambda, c_\lambda > 0$ and $0 < p_\lambda < 1$ depend on the parameter $\lambda \in [0,1)$ of the graph G.

Corollary 5. *The statement of the lemma remains true if we change A_1 from constant to drawn from any distribution.*

We prove this lemma in Section 4; the proof is a variation on the ideas from Artstein-Avidan and Milman [3].

Now we can reformulate our main result.

Theorem 6. *Let A_1 and A_2 be constructed as above (A_1 has $\Theta(\log n)$ independent entries, the rows of A_2 come from a random walk on an expander); let $A = A_1 \bullet A_2$, $E = Ker\, A$. Then, with probability $1 - o(1)$,*

$$\frac{c'\xi}{\sqrt{\log 1/\xi}} \sqrt{N}\|x\|_2 \leq \|x\|_1 \leq \sqrt{N}\|x\|_2 \quad \textit{for every} \quad x \in E \ , \tag{3}$$

where $c' > 0$ is a universal constant.

The proof uses the Lemmata formulated above as well as the following standard lemma from asymptotic convex geometry.

Lemma B. Let A be a random $n \times N$ sign matrix such that:

1. $\mathbb{P}[\|A\| > 3\sqrt{N}] \leq q$;
2. There exist $0 < p < 1$, $\varepsilon > 0$ and $C > 0$ such that for every $y \in \mathbb{R}^N$,
$$\mathbb{P}\left\{\|Ay\|_2 < 6\varepsilon\sqrt{N}\|y\|_2\right\} < Cp^n .$$

Then with probability at least

$$1 - q - p^{\Theta(n)}$$

over the choice of A, we have:

$$\|x\|_1 \geq \delta\sqrt{N}\|x\|_2 \quad \text{for every} \quad x \in \mathrm{Ker} A ,$$

where we can take

$$\delta = \frac{c\varepsilon}{\sqrt{\frac{1}{\xi}\log\frac{1}{p}\log\left(\frac{1}{\xi}\log\frac{1}{p}\right)}} ,$$

$c > 0$ being a universal constant.

For completeness, we prove Lemma B in Appendix B.

Proof of Theorem 6. According to Corollary 3 the random matrix A satisfies the condition 1. of Lemma B with $q = 1/n$. According to Corollary 5 A also satisfies 2., with $p = p_\lambda$, $C = C_\lambda$ and $\varepsilon = c_\lambda\sqrt{\xi}$. Now apply Lemma B; note that $\lambda \leq 0.95 < 1$ is bounded away from 1 and hence p_λ and C_λ may be replaced by universal constants ($p_{0.95}$ and $C_{0.95}$, resp.) □

Clearly, Theorem 6 implies Theorem 1.

3 Operator norm of a matrix with $2k$-wise independent entries

Proof of Lemma 2. We start by bounding the expectation of $\|V\|^{2k}$. For a real symmetric $n \times n$ matrix W, denote by $\lambda_1(W), \cdots, \lambda_n(W)$ the eigenvalues of W, and let $\lambda_{\max}(W) = \max_i \lambda_i(W)$. Observe that

$$\|V\|^2 = \lambda_{\max}(V^tV) = \lambda_{\max}(VV^t)$$

and hence:

$$\begin{aligned}\mathbb{E}\|V/\sqrt{N}\|^{2k} &= \mathbb{E}\lambda_{\max}(VV^t/N)^k \\ &\leq \mathbb{E}\sum_{i=1}^{n} \lambda_i(VV^t/N)^k = \mathbb{E}\mathrm{Tr}((VV^t/N)^k)\ .\end{aligned}$$

The trace of $(VV^t)^k$ is equal to

$$\sum V_{i_1,j_1}V_{i_2,j_1}V_{i_2,j_2}V_{i_3,j_2}\cdots V_{i_k,j_k}V_{i_1,j_k}\ ,$$

where the sum is over closed paths $(i_1, j_1, ..., i_k, j_k, i_1)$ in the bipartite graph $K_{n,N}$. The expectation of each term in the sum is 0 if there is some $V_{i,j}$ that appears an odd number of times, and 1 if all the terms appear an even number of times. So, the expectation is equal to the number $m(k; n, N)$ of closed even paths of length $2k$ in $K_{n,N}$, starting on the side of size n (an even path is a path in which every edge appears an even number of times).

Instead of estimating this expectation directly, we follow an idea of Aubrun [5] and take a different route. The trace of $(VV^t)^k$ is a sum over products of powers of at most $2k$ elements from V, and so, since the elements of V come from a $2k$-wise independent probability space, the expectation is the same as if the elements of V were truly independent. Hence, we can use estimates known for matrices with i.i.d. elements.

We chose to use such an estimate for matrices with Gaussian i.i.d elements. Let $\widetilde{V}$ be an $n \times N$ matrix, whose entries are independent, $\widetilde{V}_{i,j} \sim N(0,1)$. For every entry $1 \leq i \leq n$, $1 \leq j \leq N$ and every integer $l \geq 1$ we have:

$$\mathbb{E}\widetilde{V}_{i,j}^{2l} \geq (\mathbb{E}\widetilde{V}_{i,j}^2)^l = 1 = \mathbb{E}V_{i,j}^{2l}\ ; \quad \mathbb{E}\widetilde{V}_{i,j}^{2l+1} = 0 = \mathbb{E}V_{i,j}^{2l+1}\ .$$

Therefore

$$\mathbb{E}\mathrm{Tr}((VV^t/N)^k) \le \mathbb{E}\mathrm{Tr}((\widetilde{V}\widetilde{V}^t/N)^k) = \mathbb{E}\sum_{i=1}^n \lambda_i(\widetilde{V}\widetilde{V}^t/N)^k$$
$$\le n\mathbb{E}\lambda_{\max}(\widetilde{V}\widetilde{V}^t/N)^k = n\,\mathbb{E}\|\widetilde{V}/\sqrt{N}\|^{2k} .$$

We use the following bound for Gaussian random matrices with independent entries (see Davidson–Szarek [6, Thm. II.13], extending an idea of Y. Gordon):

$$\mathbb{P}\left\{\|\widetilde{V}/\sqrt{N}\| \ge 1 + \sqrt{\xi} + t\right\} < \exp(-Nt^2/2) , \quad t \ge 0 .$$

Now,

$$\mathbb{E}\|\widetilde{V}/\sqrt{N}\|^{2k} = \int_0^\infty 2kt^{2k-1}\mathbb{P}\left\{\|\widetilde{V}/\sqrt{N}\| \ge t\right\} dt$$
$$< (1+\sqrt{\xi})^{2k} + 2k\int_0^\infty (1+\sqrt{\xi}+u)^{2k-1}\exp(-Nu^2/2)\,du .$$

It is easy to see that the second term is smaller than the first one:

$$2k\int_0^\infty (1+\sqrt{\xi}+u)^{2k-1}\exp(-Nu^2/2)\,du$$
$$< 2k(1+\sqrt{\xi})^{2k-1}\int_0^\infty \exp\left\{\frac{2k-1}{1+\sqrt{\xi}}\,u - Nu^2/2\right\} du$$
$$< \frac{2k}{\sqrt{N}}(1+\sqrt{\xi})^{2k-1}\int_{-\infty}^\infty \exp\left\{\frac{2k-1}{\sqrt{N}+\sqrt{n}}\,u - u^2/2\right\} du$$
$$= (1+\sqrt{\xi})^{2k-1}\frac{\sqrt{8\pi}\,k}{\sqrt{N}}\exp\left\{\frac{1}{2}\left(\frac{2k-1}{\sqrt{N}+\sqrt{n}}\right)^2\right\}$$
$$= (1+\sqrt{\xi})^{2k} \times O(k/\sqrt{N}) \times e^{O(k^2/N)} .$$

If $k \le c_2\sqrt{N}$ (for an appropriately chosen numerical constant $c_2 > 0$), the product of the O-terms is not greater than 1. Hence

$$\mathbb{E}\|\widetilde{V}/\sqrt{N}\|^{2k} < 2(1+\sqrt{\xi})^{2k} ,$$

implying that

$$\mathbb{E}\|V/\sqrt{N}\|^{2k} < 2n(1+\sqrt{\xi})^{2k} .$$

Now by Chebyshev's inequality

$$\mathbb{P}\left\{\|V/\sqrt{N}\| \ge 1+\sqrt{\xi}+t\right\} \le \frac{\mathbb{E}\|V/\sqrt{N}\|^{2k}}{(1+\sqrt{\xi}+t)^{2k}} < 2n\left(\frac{1+\sqrt{\xi}}{1+\sqrt{\xi}+t}\right)^{2k}$$

□

Remarks.

1. The lemma shows that for $k = \Omega(\log N)$ the operator norm of $V/\sqrt{N}$ is not much larger than $1 + \sqrt{\xi}$. This matches the bound for matrices with independent entries (cf. Geman [8]).
2. A more direct proof would be to bound the numbers $m(k; n, N)$ directly, as in the work of Geman [8]. This would yield an estimate similar to the one we get.

4 Bound for a single vector

Fix x, $\|x\|_2 = 1$; let us bound the probability

$$\mathbb{P}\left\{\|Ax\|_2 < 6\varepsilon\sqrt{N}\right\}$$

when $A = A_1 \bullet A_2$, A_1 is a fixed sign matrix and A_2 is generated from a random walk on an expander as explained in Section 2.

Recall that $G = (\mathcal{V}, \mathcal{E})$ is a d-regular graph with $2^{O(\log N)}$ vertices, and P^G is the transition matrix of the random walk on G; λ is the second largest absolute value of an eigenvalue of P^G.

First we bound from below the probability that a coordinate of Ax is not very small.

Lemma 7. *Let Ψ be a random vector in $\{-1, +1\}^N$ with 4-wise independent coordinates. Then*

$$\mathbb{P}\left\{\langle \Psi, x\rangle^2 \geq 1/2\right\} \geq 1/12 \ .$$

Proof. First,

$$\mathbb{E}\langle \Psi,\, x\rangle^2 = \sum_{i,j=1}^{N} x_i x_j \mathbb{E}\Psi_i\Psi_j = \sum_{i=1}^{N} x_i^2 = 1 \ ;$$

$$\begin{aligned}\mathbb{E}\langle \Psi,\, x\rangle^4 &= \sum_{i,j,k,l=1}^{N} x_i x_j x_k x_l \, \mathbb{E}\Psi_i\Psi_j\Psi_k\Psi_l \\ &= \sum_{i=1}^{N} x_i^4 + 6 \sum_{1\leq i<j\leq N} x_i^2 x_j^2 < 3\left(\sum_{i=1}^{N} x_i^2\right)^2 = 3 \ .\end{aligned}$$

Recall the Paley–Zygmund inequality [14]:

Lemma (Paley–Zygmund)**.** *If $Z \geq 0$ is a random variable with finite second moment, $0 < \theta < 1$, then*

$$\mathbb{P}\{Z \geq \theta \mathbb{E} Z\} \geq (1-\theta)^2 \frac{\mathbb{E}(Z)^2}{\mathbb{E}(Z^2)} \ .$$

Applying the inequality for $Z = \langle \Psi, x \rangle^2$, $\theta = 1/2$, we obtain the statement of the lemma.

□

Proof of Lemma 4. Let us show that a constant fraction of the rows ψ_i of A satisfy w.h.p

$$\langle \psi_i, x \rangle \geq 1/2 \ . \tag{4}$$

For fixed A_1 and $1 \leq i \leq n$, the coordinates of ψ_i are 4-wise independent; therefore by Lemma 7 there is a subset $S_i \subset \mathcal{V}$ such that $|S_i|/|\mathcal{V}| \geq 1/12$, and the i-th ψ_i of A satisfies (4) iff the i-th row v_i of A_2 lies in S_i.

We need a modification of Kahale's Chernoff-type bound on expanders [12], see also Alon, Feige, Wigderson and Zuckerman [2, Theorem 4], Artstein-Avidan and Milman [3, Section 4], and Hoory, Linial and Wigderson [9, Theorem 3.11] for related results[1]

Lemma 8. *Let $G = (\mathcal{V}, \mathcal{E})$ be a graph; as before, let $1 = \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots$ be the eigenvalues of P^G; denote $\lambda = \max_{i \geq 2} |\lambda_i|$. The probability that a random walk on G, starting from a random point in $\mathcal{V}$, is in S_i on the i-th step, $i = 1, 2, \cdots, k$, is at most*

$$\prod_{i=1}^{k-1} \sqrt{\lambda + (1-\lambda)\frac{|S_i|}{|V|}} \sqrt{\lambda + (1-\lambda)\frac{|S_{i+1}|}{|V|}} \ .$$

Proof of Lemma 8. Denote $e = (1, 1, \cdots, 1)/\sqrt{|\mathcal{V}|}$, and denote by Π_i the projector on the coordinates in S_i. Then the probability in question equals

$$\begin{aligned} &\langle \Pi_k P^G \Pi_{k-1} P^G \cdots P^G \Pi_1 e, e \rangle \\ &\qquad \leq \|\Pi_k P^G \Pi_{k-1}\| \times \|\Pi_{k-1} P^G \Pi_{k-2}\| \times \cdots \times \|\Pi_2 P^G \Pi_1\| \ , \end{aligned} \tag{5}$$

[1]Added in proof: an even stronger result was recently proved. See theorem 5.4 in E. Mossel, R. O'Donnell, O. Regev, J. Steif and B. Sudakov, Non-Interactive Correlation Distillation, Inhomogeneous Markov Chains and the Reverse Bonami-Beckner Inequality, Israel Journal of Mathematics 154 (2006), 299-336.

where we used the submultiplicativity of operator norm and the equality $\Pi_i^2 = \Pi_i$. Let us bound the norms

$$\|\Pi_{i+1} P^G \Pi_i\| = \max_{\|g\|_2=1} \|\Pi_{i+1} P^G \Pi_i g\|_2 \ .$$

First of all, the vector g for which the maximum is attained is supported in S_i; hence $\Pi_i g = g$. Let us decompose $g = \alpha e + \beta v$, where $\alpha^2 + \beta^2 = 1$ and v is a unit vector orthogonal to e.

Note that

$$|\alpha| = |\langle g, e\rangle| \le \|g\|_1 / \sqrt{|\mathcal{V}|} \le \sqrt{\frac{|S_i|}{|\mathcal{V}|}} \, \|g\|_2 = \sqrt{\frac{|S_i|}{|\mathcal{V}|}} \ .$$

Therefore $P^G g = \alpha e + \beta P^G v$. Now,

$$\|\Pi_{i+1} P^G g\|_2 = \max_{\|h\|_2=1} \langle \Pi_{i+1} P^G g, h\rangle = \max_{\|h\|_2=1} \langle P^G g, \Pi_{i+1} h\rangle \ ;$$

we may assume that h is supported in S_{i+1}. Let $h = \alpha' e + \beta' v'$, where v' is a unit vector orthogonal to e; as before,

$$\alpha'^2 + \beta'^2 = 1 \quad \text{and} \quad |\alpha'| \le \sqrt{\frac{|S_{i+1}|}{|\mathcal{V}|}} \ .$$

Hence

$$\begin{aligned}
\langle P^G g, h\rangle &= \alpha\alpha' + \beta\beta' \langle P^G v, v'\rangle \le \alpha\alpha' + \lambda\beta\beta' \\
&\le \sqrt{\alpha^2 + \lambda\beta^2} \, \sqrt{\alpha'^2 + \lambda\beta'^2} \\
&= \sqrt{\lambda + (1-\lambda)\alpha^2} \, \sqrt{\lambda + (1-\lambda)\alpha'^2} \\
&\le \sqrt{\lambda + (1-\lambda)\frac{|S_i|}{|\mathcal{V}|}} \, \sqrt{\lambda + (1-\lambda)\frac{|S_{i+1}|}{|\mathcal{V}|}} \ .
\end{aligned}$$

□

Now, if $\|Ax\|_2 < 6\varepsilon\sqrt{N}$, A has at most $72\varepsilon^2 N$ rows ψ such that

$$\langle \psi, x\rangle^2 \ge 1/2 \ .$$

By Lemma 8, the probability of this event is at most

$$\begin{aligned}
&\binom{n}{[72\varepsilon^2 N]} \left(\frac{11}{12}(1-\lambda) + \lambda\right)^{n - [72\varepsilon^2 N] - 1} \\
&\qquad \le 2\left(\frac{e\xi}{72\varepsilon^2}\right)^{72n\varepsilon^2/\xi} \left(\frac{11}{12}(1-\lambda) + \lambda\right)^{n - 72n\varepsilon^2/\xi} \ . \qquad (6)
\end{aligned}$$

For ε small enough, this probability is exponentially small. More formally, it is easy to see that there exist some constants $C_\lambda \geq 1 > c_\lambda > 0$ and $0 < p_\lambda < 1$ depending only on λ, such that

$$\mathbb{P}\left\{\|Ax\|_2 < 6\varepsilon\sqrt{N}\right\} \leq C_\lambda p_\lambda^n \quad \text{if} \quad 0 < \varepsilon \leq c_\lambda\sqrt{\xi}\ . \tag{7}$$

Lemma 4 is proved.

□

A Construction of k-wise independent random bits

For completeness, we recall the construction of $2^r - 1$ k-wise independent random bits from kr independent random bits due to Alon, Babai and Itai [1]. It will be more convenient to work with vectors of $\{0,1\}$ rather than $\{-1,+1\}$.

Let

$$\alpha_1, \cdots, \alpha_{2^r-1} \in \mathrm{GF}(2^r)$$

be the non-zero elements of the finite field of cardinality 2^r. $\mathrm{GF}(2^r)$ is a linear space over $\mathrm{GF}(2)$; hence we may represent an element $\alpha \in \mathrm{GF}(2^r)$ as an r-tuple $\widetilde{\alpha} \in \mathrm{GF}(2)^r$.

Consider the matrix

$$M = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{k-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{k-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \alpha_{2^r-1} & \alpha_2^2 & \cdots & \alpha_{2^r-1}^{k-1} \end{pmatrix}.$$

Every k rows of M form a Van der Monde matrix, and in particular are linearly independent. Let

$$\widetilde{M} = \begin{pmatrix} 1 & \widetilde{\alpha}_1 & \widetilde{\alpha}_1^2 & \cdots & \widetilde{\alpha}_1^{k-1} \\ 1 & \widetilde{\alpha}_2 & \widetilde{\alpha}_2^2 & \cdots & \widetilde{\alpha}_2^{k-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \widetilde{\alpha}_{2^r-1} & \widetilde{\alpha}_2^2 & \cdots & \widetilde{\alpha}_{2^r-1}^{k-1} \end{pmatrix}$$

be the corresponding $kr \times (2^r - 1)$ matrix over $\mathrm{GF}(2)$; its rows are also linearly independent. Now let Z be a random vector distributed uniformly in $\mathrm{GF}(2)^{kr}$; let $X = \widetilde{M}Z$.

Claim. The coordinates of the vector X are k-wise independent.

Proof. For every set of indices $\emptyset \neq I \subset \{1, \cdots, 2^r - 1\}$ such that $|I| = k$, the matrix $\widetilde{M}_I$ formed from the corresponding rows of $\widetilde{M}$ is of rank k; that is, $\widetilde{M}_I$ is surjective and the preimages of the vectors in $\{0,1\}^k$ are of equal size. The vector Z is distributed uniformly in $\mathrm{GF}(2)^{kr}$; hence the vector $(X_i)_{i \in I} = \widetilde{M}_I Z$ is uniformly distributed in $\mathrm{GF}(2)^k$. □

B Proof of Lemma B

The proof of Lemma B is based on ε-net arguments.

Definition 5. Let $S \subset \mathbb{R}^N$ be a convex set. A (finite) subset $\mathcal{N} \subset S$ is called an ε-net in S if for every $x \in S$ there exists $y \in \mathcal{N}$ such that $\|x - y\|_2 \leq \varepsilon$.

Notation. Let $t > 0$ and let $K \subset \mathbb{R}^n$ be a convex body. As usual, denote

$$tK = \{tx \,|\, x \in K\} \ .$$

Similarly to [3], we use the following result, due to Schütt [15]:

Theorem (Schütt)**.** *The exists a universal constant $c > 0$ such that for any $\zeta > 0$ and $\theta \geq c\sqrt{\frac{1}{\zeta}\log\frac{1}{\zeta}}$ there exists a θ-net $\mathcal{N}$ in $\sqrt{N}B_1^N$ such that $|\mathcal{N}| \leq e^{\zeta N}$.*

Proof of Lemma B. Pick $0 < \zeta < \xi \log\frac{1}{p}$; then $e^\zeta < 1/p^\xi$. Set

$$\delta = \frac{\varepsilon}{c\sqrt{\frac{1}{\zeta}\log\frac{1}{\zeta}}} \ .$$

Scaling the result of Schütt's theorem times δ, we get an ε-net $\mathcal{N}$ in $\delta\sqrt{N}B_1^N$, $|\mathcal{N}| \leq e^{\zeta N}$.

By our assumptions, for every $y \in \mathcal{N}$

$$\mathbb{P}\left\{\|Ay\|_2 < 6\varepsilon\sqrt{N}\|y\|_2\right\} < Cp^n \ ,$$

and so the probability that there exists $y \in \mathcal{N}$ with

$$\|Ay\|_2 < 6\varepsilon\sqrt{N}\|y\|_2$$

is at most

$$Ce^{\zeta N}p^n = p^{\Theta(n)} \ .$$

Assume that for every $y \in \mathcal{N}$ we have

$$\|Ay\|_2 \geq 6\varepsilon\sqrt{N}\|y\|_2 \ ,$$

and also that $\|A\| \leq 3\sqrt{N}$. This event happens with probability at least $1 - q - p^{\Theta(n)}$. We will show that whenever these two conditions hold, every $x \in \mathrm{Ker}A$ satisfies

$$\|x\|_1 \geq \delta\sqrt{N}\|x\|_2 \ .$$

It is enough to show this for x with $\|x\|_2 = 1$.

Take any $x \in \mathbb{R}^N$ with $\|x\|_1 < \delta\sqrt{N}$ and $\|x\|_2 = 1$. We will show $x \notin \mathrm{Ker}(A)$. First, $x \in \delta\sqrt{N}B_1^N$, and so there exists $y \in \mathcal{N}$ such that $\|x - y\| \leq \varepsilon$. Now we have:

$$\begin{aligned}\|Ax\|_2 &\geq \|Ay\|_2 - \|A(x-y)\|_2 \geq 6\varepsilon\sqrt{N}\|y\|_2 - \|A\|\|x-y\|_2 \\ &\geq 6\varepsilon(1-\varepsilon)\sqrt{N} - 3\varepsilon\sqrt{N} > 0 \ ,\end{aligned}$$

where we used the fact that

$$\|y\|_2 \geq \|x\|_2 - \|x - y\|_2 \geq 1 - \varepsilon \ .$$

□

References

[1] N. Alon, L. Babai and A. Itai, A fast and simple randomized parallel algorithm for the maximal independent set problem, J. Algorithms 7 (1986), no. 4, 567–583

[2] N. Alon, U. Feige, A. Wigderson and D. Zuckerman, Derandomized graph products, Comput. Complexity 5 (1995), no. 1, 60–75

[3] S. Artstein-Avidan and V. D. Milman, Logarithmic reduction of the level of randomness in some probabilistic geometric constructions, J. Funct. Anal. 235 (2006), no. 1, 297–329

[4] S. Artstein-Avidan and V. D. Milman, Using Rademcher Permutations to reduce randomness, *to appear*

[5] G. Aubrun, Sampling convex bodies: a random matrix approach, *to appear*

[6] K. R. Davidson, S. J. Szarek, Local operator theory, random matrices and Banach spaces, Handbook of the geometry of Banach spaces, Vol. I, 317–366, North-Holland, Amsterdam, 2001

[7] T. Figiel, J. Lindenstrauss and V. D. Milman, The dimension of almost spherical sections of convex bodies, Acta Math. 139 (1977), no. 1-2, 53–94

[8] S. Geman, A limit theorem for the norm of random matrices, Ann. Probab. 8 (1980), no. 2, 252–261

[9] S. Hoory, N. Linial and A. Wigderson, Expander graphs and their applications, Bull. Amer. Math. Soc. (N.S.) 43 (2006), no. 4, 439–561 (electronic)

[10] P. Indyk, Uncertainty Principles, Extractors, and Explicit Embeddings of L_2 into L_1, ECCC TR06-126, 2006

[11] P. Indyk, Stable distributions, pseudorandom generators, embeddings and data stream computation, 41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000), 189–197, IEEE Comput. Soc. Press, Los Alamitos, CA, 2000

[12] N. Kahale, Better expansion for Ramanujan graphs, 32nd Annual Symposium on Foundations of Computer Science (San Juan, PR, 1991), 398–404, IEEE Comput. Soc. Press, Los Alamitos, CA, 1991

[13] B. S. Kashin, The widths of certain finite-dimensional sets and classes of smooth functions, Izv. Akad. Nauk SSSR Ser. Mat. 41 (1977), no. 2, 334–351, 478

[14] R. E. A. C. Paley and A. Zygmund, A note on analytic functions in the unit circle, Proc. Camb. Phil. Soc. 28, 1932, 266-272

[15] C. Schütt, Entropy numbers of diagonal operators between symmetric Banach spaces., J. Approx. Theory 40 (1984), no. 2, 121–128

[16] S. Szarek, Convexity, Complexity, and High Dimensions, *to appear in* Proceedings of the International Congress of Mathematicians (Madrid, 2006)

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