Electronic Colloquium on Computational Complexity, Report No. 194 (2018)

ANTI-CONCENTRATION IN MOST DIRECTIONS

AMIR YEHUDAYOFF

ABSTRACT. We prove anti-concentration for the inner product of two independent random vectors in the discrete cube. Our results imply Chakrabarti and Regev's lower bound on the randomized communication complexity of the gap-hamming problem. They are also meaningful in the context of randomness extraction. The proof provides a framework for establishing anti-concentration in discrete domains. The argument has two different components. A local component that uses harmonic analysis, and a global ("information theoretic") component.

1. INTRODUCTION

Anti-concentration of a random processes means that the distribution of outcomes is not concentrated in a small region. No single outcome is obtained too often. It plays an important role in mathematics and computer science (see e.g. [11, 16, 1] and references within).

The standard example is a sum of i.i.d. random variables. If X in $\mathcal{X} = \{\pm 1\}^n$ is fixed, and B is uniformly distributed in $\mathcal{X}$, then the random integer $\langle B, X\rangle = \sum_i B_i X_i$ is anti-concentrated. The probability that $\langle B, X\rangle$ takes any specific value is at most $O(1/\sqrt{n})$. This was studied and generalized by Littlewood and Offord [11], Erdös [7], and many others. Higher dimensional analogs of this phenomenon were studied by Frankl and Furedi [8], Halász [9] and others.

It is interesting to understand the generality of this phenomenon (see also [15] and references within). Anti-concentration certainly fails when the entropy of B is not full. We can, for example, condition B on the pretty likely event that $\langle B, X\rangle = 0$.

Can we somehow recover anti-concentration? A natural suggestion is to allow X to be random as well. This indeed recovers anti-concentration, as the following theorem shows.

Theorem (Chakrabarti and Regev [4]). *There is a constant $c > 0$ so that the following holds. Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{X}$ be of sizes at least $2^{(1-c)n}$. If*

Partially supported by ISF grant 1162/15. This work was done while the author was visiting the Simons Institute for the Theory of Computing.

ISSN 1433-8092

(A,B) is uniformly distributed in $\mathcal{A}\times\mathcal{B}$ then

$$\mathbb{P}[|\langle A,B\rangle| \leq c\sqrt{n}] \leq 1-c.$$

Chakrabarti and Regev's proof uses the deep connection between the discrete cube and gaussian space. They proved a geometric correlation inequality in gaussian space, and translated it to the cube. Vidick [18] later simplified part of their argument, but stayed in the geometric setting. Sherstov [13] found a third proof that uses Talagrand's inequality from convex geometry [14] and ideas of Babai, Frankl and Simon from communication complexity [2].

We generalize the theorem above.

Theorem 1. *For every $\beta > 0$, there are $c, C > 0$ so that the following holds. Let $\mathcal{A},\mathcal{B} \subseteq \mathcal{X}$ be so that $|\mathcal{B}| = 2^{\beta n}$ and $|\mathcal{A}| \geq 2^{(1-c)n}$. If (A,B) is uniformly distributed in $\mathcal{A}\times\mathcal{B}$ then for all $I \subset \mathbb{Z}$,*

$$\mathbb{P}[\langle A,B\rangle \in I] \leq C\big(\tfrac{|I|^2}{n}\big)^{1/4}.$$

Theorem 1 directly implies Chakrabarti and Regev's theorem. The measure of any set of size much smaller than $\sqrt{n}$ is bounded away from 1. This holds even when $\mathcal{B}$ is quite small, say $|\mathcal{B}| = 2^{n/10}$. The theorem also implies a point-wise bound: $\mathbb{P}[\langle A,B\rangle = z] = O(n^{-1/4})$ for all $z \in \mathbb{Z}$ (below we provide a stronger point-wise bound).

When studying anti-concentration, what we are ultimately interested in is proving point-wise estimates. Namely, we would like to control the *concentration probability* or the ℓ_∞ norm[1]

$$\|\nu\|_\infty = \max_{\omega\in\Omega} \nu(\omega)$$

(see [16] and references within). Although it is the strongest measure of anti-concentration, it is not analytic. Other norms are, therefore, often more convenient to work with.

The ℓ_r norm is defined as $\|\nu\|_r = \big(\sum_\omega (\nu(\omega))^r\big)^{1/r}$. The corresponding Rényi entropy of ν is $H_r(\nu) = \frac{r}{1-r}\log\|\nu\|_r$ for $r > 1$. The norm and the entropy are inversely related; the smaller the norm, the larger the entropy, and vice versa. The norm yields the following type of anti-concentration.

Claim 2. *For every $r \geq 1$ and event I, we have $\nu(I) \leq |I|^{(r-1)/r}\|\nu\|_r$.*

Our main result is a general upper bound on the norms of the distribution of interest. Let X be uniformly distributed in $\mathcal{X}$. Let $\mathcal{B}$ be a family of vectors in $\mathcal{X}$ of size $2^{\beta n}$. Let μ_X be the distribution of $\langle B,X\rangle$ with X fixed and B uniformly distributed in $\mathcal{B}$.

[1] We consider only finite probability spaces in this text.

Theorem 3. *For every $\beta > 0$, there are $c, C > 0$ so that for each $r \geq 2$,*

$$\mathop{\mathbb{P}}_{X}\left[\|\mu_X\|_r > C(\tfrac{r}{n^{(r-1)/r}})^{1/2}\right] < C2^{-cn}.$$

The theorem is sharp in the sense that even when $\mathcal{B} = \mathcal{X}$, the norm $\|\mu_X\|_r$ is roughly $(\frac{1}{n^{(r-1)/(r)}})^{1/2}$. There is a gap of order $r^{1/2}$ between the upper and lower bounds.

The ℓ_∞ norm can be approximated by the ℓ_r norm for large values of r. Taking $r \approx \log n$ we get the following estimate on the concentration probability that is sharp up to a factor of order $(\log n)^{1/2}$.

Corollary 4. *For every $\beta > 0$, there are $c, C > 0$ so that*

$$\mathop{\mathbb{P}}_{X}\left[\|\mu_X\|_\infty > C(\tfrac{\log n}{n})^{1/2}\right] < C2^{-cn}.$$

The corollary allows to strengthen Theorem 1 for singletons and small sets. Let ν denote by distribution of $\langle A, B\rangle$ from Theorem 1. The theorem implies that $\|\nu\|_\infty$ is at most order $(\frac{1}{n})^{1/4}$. Corollary 4 implies a stronger estimate: $\|\nu\|_\infty$ is at most order $(\frac{\log n}{n})^{1/2}$. The true value should be order $(\frac{1}{n})^{1/2}$. This remains open.

To prove the results above, we build a framework for proving anti-concentration results in discrete domains. Think of the random variable $\langle B, X\rangle$ as built in n steps. It starts as 0, and $B_d X_d$ is added to $\langle B_{<d}, X_{<d}\rangle$ to generate $\langle B_{\leq d}, X_{\leq d}\rangle$. To analyze the behavior of this system, we first show that locally (in the microscopic scale) entropy often increasing (Section 2). This part of the argument uses harmonic analysis (even though our ultimate goal is not the ℓ_2 norm). The second part of the argument is macroscopic (Section 3). We identify a global event that guarantees that the small local increments in entropy yield substantial entropy in the whole system. The last step is proving that the macroscopic event almost always holds. This is achieved by an encoding argument. Situations where the macroscopic entropy is not high can be described by a small number of bits.

There are several differences between our argument and the ones in [4, 18, 13]. The main difference is that the arguments from [4, 18, 13] are based, in one way or another, on the geometry of euclidean space. The arguments in [4, 18] prove a correlation inequality in gaussian space and translate it to the discrete world. It seems that such an argument can not yield effective bounds on the concentration probability in the discrete setting. A common ingredient to [4, 13] is showing that every set of large enough measure contains many almost orthogonal vectors (this is called "identifying the hard core" in [13]). In [18] this part of the argument is replaced by a statement about a relevant matrix. Our argument does not contain any such step.

Related topics.

Communication complexity. Chakrabarti and Regev's main motivation was understanding the randomized communication complexity of the gap-hamming problem. The gap-hamming problem was introduced by Indyk and Woodruff in the context of streaming [10]. Proving lower bounds on its communication complexity was a central open problem for almost ten years, until Chakrabarti and Regev solved it [4]. Vidick [18] and Sherstov [13] later simplified the proof.

Theorem 1 also implies the lower bound for the randomized communication complexity of the gap-hamming problem (see e.g. [13]). As opposed to [4, 18, 13], the proof presented here lies entirely in the discrete domain. The underlying ideas may therefore be of independent interest.

Pseudorandomness. Randomness is a computational resource [17]. There are many sources of randomness, and some of them are *weak* or imperfect. Randomness extractors allow to use weak sources of randomness as if they were perfect.

The study of randomness extractors is about constructing explicit maps that transform weak sources of randomness to almost uniform outputs. The main goal is generating a uniform output in the most general scenario possible. This often requires ingenious constructions.

The scenario described above fits nicely in the context of *two-source extractors*. A two-source extractor maps two independent random variables A and B with significant min-entropy to a single almost uniform output.

Chor and Goldreich [6] used Lindsey's lemma to show that inner product modulo two is a two-source extractor. The bit $\langle A, B\rangle \ \mathsf{mod}\ 2$ is close to a uniform random bit as long as $|\mathcal{A}| \cdot |\mathcal{B}| \gg 2^n$. Bourgain [3], Raz [12] and Chattopadhyay and Zuckerman [5] constructed two-source extractors with much better parameters.

This work can be interpreted as studying a related but somewhat different question. The high-level suggestion is to investigate what other pseudorandom properties known extractors satisfy.

We already know that inner product is an excellent two-source extractor. Now we also know that over the integers inner product is anti-concentrated. This is not as good as being uniform, but inner product is not uniform over the integers (it is binomial).

Remark. *The way $\frac{1}{\sqrt{n}}$ emerges in the proof is quite surprising. As an example, consider the sequence recursively defined by $a_0 = 1$ and $a_{d+1} = a_d(1 - 0.5a_d^2)$. The observation is that $a_n = O(\frac{1}{\sqrt{n}})$. This* *recursion comes from harmonic analysis (Lemma 6). The second power of a_d in the recursion comes from the second power in the approximation* $\cos(\xi) \approx 1 - \xi^2$.

2. Microscopically

Here we analyze the local increases in entropy. The argument is spectral and uses harmonic analysis. We work over the abelian group $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$ with $N = 2n+1$. The choice of N ensures that there is a one-to-one correspondence between the integer $\langle B, X \rangle$ and the group element $\langle B, X \rangle \bmod N$.

2.1. Harmonic analysis.

2.1. Harmonic analysis. The group $\mathbb{Z}_N$ acts on the vector space of functions from $\mathbb{Z}_N$ to $\mathbb{C}$. This vector space is endowed with the standard inner product $\langle f, g \rangle = \sum_{z \in \mathbb{Z}_N} f(z)\overline{g(z)}$ where $\overline{\xi}$ is the complex conjugate of $\xi \in \mathbb{C}$. For $z \in \mathbb{Z}_n$, let S_z be the operator that shifts the function $f : \mathbb{Z}_N \to \mathbb{C}$ by z. That is, $S_z f(x) = f(x - z)$ for all x. The shifts are unitary and they commute. Let $\{e_z : z \in \mathbb{Z}_N\}$ be the set of the N normalized eigenvectors:

$$e_z(x) = \frac{e^{2\pi i \frac{zx}{N}}}{\sqrt{N}}.$$

The eigenvalue of e_z with respect to the shift S_1 is $\lambda_z = e^{-2\pi i \frac{z}{N}}$. The Fourier transform of f is $\widehat{f} : \mathbb{Z}_N \to \mathbb{C}$ defined by

$$\widehat{f}(z) = \langle f, e_z \rangle .$$

Remark. *Harmonic analysis naturally allows to work with the ℓ_2 norm. Our goal is to analyze general ℓ_r norms. We, therefore, need to translate the problem from ℓ_r to ℓ_2.*

2.2. Entropy locally increases. The first observation is that entropy does not decrease (the second law of thermodynamics). Assume we have two distribution μ_1 and μ_{-1} on $\mathbb{Z}_N$. Think about them as distributions of two particles Z_1 and Z_{-1}. We use them to generate a new distribution

$$\mu = \gamma S_1 \mu_1 + (1 - \gamma) S_{-1} \mu_{-1}$$

where $\gamma \in [0, 1]$. Stated differently, we consider the particle

$$Z = \begin{cases} Z_1 + 1 & \text{with probability } \gamma, \\ Z_{-1} - 1 & \text{with probability } 1 - \gamma. \end{cases}$$

The entropy of the new particle is not smaller than the average entropy of the old particles.

Observation 5. *For every $r \geq 1$, for every two distributions μ_1 and μ_{-1} on $\mathbb{Z}_N$ and for every $\gamma \in [0,1]$,*

$$\|\gamma S_1 \mu_1 + (1-\gamma) S_{-1}\mu_{-1}\|_r^r \leq \gamma \|\mu_1\|_r^r + (1-\gamma)\|\mu_{-1}\|_r^r.$$

The main lemma is a quantitative estimate on the increase in entropy when one new bit of randomness is inserted into the system (the randomness determines which shift is applied).

Lemma 6. *For every $r \geq 2$, for every two distributions μ_1 and μ_{-1} on $\mathbb{Z}_N$ and for every $\gamma \in [0,1]$,*

$$\begin{aligned}
&(1)\quad \tfrac{1}{2}(\|\gamma S_1\mu_1 + (1-\gamma)S_{-1}\mu_{-1}\|_r^r + \|\gamma S_{-1}\mu_1 + (1-\gamma)S_1\mu_{-1}\|_r^r) \\
&(2)\quad \leq \gamma\|\mu_1\|_r^r + (1-\gamma)\|\mu_{-1}\|_r^r \\
&(3)\quad\quad - \tfrac{\gamma(1-\gamma)}{120}\Big(\tfrac{\|\mu_1\|_r^{3r}}{\|\mu_1\|_{r/2}^{2r}} + \tfrac{\|\mu_{-1}\|_r^{3r}}{\|\mu_{-1}\|_{r/2}^{2r}}\Big) + \tfrac{10}{N^2}(\gamma\|\mu_1\|_r^r + (1-\gamma)\|\mu_{-1}\|_r^r).
\end{aligned}$$

The term in (1) is the entropy in the system after the new bit of entropy is introduced. The term in (2) is the entropy in the system before the new entropy arrives. The term in (3) is the increase in entropy. The $+\frac{10}{N^2}\ldots$ term in (3) is somehow necessary; if the distributions are already uniform then there is no hope for increase in entropy.

Proof of Lemma 6. Start by fixing $x \in \{\pm 1\}$, and considering

$$\mu = \gamma S_x \mu_1 + (1-\gamma) S_{-x}\mu_{-1}.$$

Move from ℓ_r to ℓ_2 by considering

$$f_1 = \mu_1^{r/2} \quad \& \quad f_{-1} = \mu_{-1}^{r/2}.$$

Convexity implies

$$\begin{aligned}
\|\mu\|_r^r &= \sum_z (|\gamma\mu_1(z-x) + (1-\gamma)\mu_{-1}(z+x)|^{r/2})^2 \\
&\leq \sum_z (\gamma|\mu_1(z-x)|^{r/2} + (1-\gamma)|\mu_{-1}(z+x)|^{r/2})^2 \\
&= \|\gamma S_x f_1 + (1-\gamma)S_{-x} f_{-1}\|_2^2.
\end{aligned}$$

Since shifts are unitary,

$$\|\mu\|_r^r \leq \gamma^2\|f_1\|_2^2 + (1-\gamma)^2\|f_{-1}\|_2^2 + 2\gamma(1-\gamma)\,\mathfrak{R}(\langle S_x f_1, S_{-x} f_{-1}\rangle),$$

where $\mathfrak{R}(\xi)$ is the real part of the complex number ξ. Focus on

$$\begin{aligned}\langle S_x f_1, S_{-x} f_{-1}\rangle &= \left\langle \sum_z \widehat{f_1}(z) S_x e_z, \sum_{z'} \widehat{f_{-1}}(z') S_{-x} e_{z'} \right\rangle \\ &= \sum_z \widehat{f_1}(z)\overline{\widehat{f_{-1}}(z)}\lambda_z \overline{\lambda_z^{-1}} \\ &= \sum_z \widehat{f_1}(z)\overline{\widehat{f_{-1}}(z)}\lambda_z^2.\end{aligned}$$

Now, average over the two options for x,

$$\begin{aligned}(1) &\leq \gamma^2\|f_1\|_2^2 + (1-\gamma)^2\|f_{-1}\|_2^2 + 2\gamma(1-\gamma)\,\mathfrak{R}\Big(\sum_{z\in Z} \widehat{f_1}(z)\overline{\widehat{f_{-1}}(z)}\tfrac{\lambda_z^2+\lambda_z^{-2}}{2}\Big) \\ &\leq \gamma^2\|f_1\|_2^2 + (1-\gamma)^2\|f_{-1}\|_2^2 + 2\gamma(1-\gamma)\sum_{z\in Z} |\widehat{f_1}(z)\widehat{f_{-1}}(z)\cos(\tfrac{4\pi z}{N})|.\end{aligned}$$

Let $\kappa : \mathbb{Z}_N \to \mathbb{R}$ be defined by $\kappa_z = \sqrt{|\cos(\frac{4\pi z}{N})|}$. Denote by $u.v$ the point-wise product of the two functions u and v. The Cauchy-Schwarz and the AM-GM inequalities imply

$$\begin{aligned}(1) &\leq \gamma^2\|f_1\|_2^2 + (1-\gamma)^2\|f_{-1}\|_2^2 + 2\gamma(1-\gamma)\|\widehat{f_1}.\kappa\|\|\widehat{f_{-1}}.\kappa\| \\ &\leq \gamma^2\|f_1\|_2^2 + (1-\gamma)^2\|f_{-1}\|_2^2 + \gamma(1-\gamma)(\|\widehat{f_1}.\kappa\|_2^2 + \|\widehat{f_{-1}}.\kappa\|_2^2).\end{aligned}$$

The following estimate is the last ingredient in the proof.

Claim 7. *For every non-zero $g : \mathbb{Z}_N \to \mathbb{C}$,*

$$\|\hat{g}.\kappa\|_2^2 \leq \Big(1 - \tfrac{\|g\|_2^4}{120\|g\|_1^4} + \tfrac{10}{N^2}\Big)\|g\|_2^2.$$

Proof. For every z,

$$|\widehat{g}(z)|^2 \leq (\|g\|_1\|e_z\|_\infty)^2 = \tfrac{\|g\|_1^2}{N}.$$

Interpreting the elements of $\mathbb{Z}_N$ as the integers $\{0,1,\dots,N-1\}$, let Z be the set of $z \in \mathbb{Z}_N$ so that the distance of $\frac{4\pi z}{N}$ from $\{0,\pi,2\pi,3\pi,4\pi\}$ is at least $\frac{\|g\|_2^2}{4\|g\|_1^2}$. There are at most $2\xi+1$ integers whose distance from 0 is at most ξ. The number of elements not in Z is hence at most

$$5 + 8\cdot 2\tfrac{N\|g\|_2^2}{16\pi\|g\|_1^2}.$$

If $\frac{5\|g\|_1^2}{N} > \frac{\|g\|_2^2}{6}$ then $\frac{10}{N^2} > \frac{\|g\|_2^4}{120\|g\|_1^4}$ and the claim holds. Otherwise,

$$\|\hat{g}\|_2^2 \leq \tfrac{\|g\|_1^2}{N}\cdot\Big(5 + \tfrac{N\|g\|_2^2}{\pi\|g\|_1^2}\Big) + \sum_{z\in Z}|\widehat{g}(z)|^2 \leq \tfrac{\|g\|_2^2}{2} + \sum_{z\in Z}|\widehat{g}(z)|^2.$$

Using the approximation $|\cos(\xi)| \le 1 - \Omega(\xi^2)$ that is valid near 0,

$$\begin{aligned} \|\hat{g}.\kappa\|_2^2 &= \sum_z |\widehat{g}(z)|^2 |\cos(\tfrac{4\pi z}{N})| \\ &\le \sum_{z \notin Z} |\widehat{g}(z)|^2 + \sum_{z \in Z} |\widehat{g}(z)|^2 |\cos(\tfrac{4\pi z}{N})| \\ &\le \sum_{z \notin Z} |\widehat{g}(z)|^2 + (1 - \tfrac{\|g\|_2^4}{60\|g\|_1^4}) \sum_{z \in Z} |\widehat{g}(z)|^2 \\ &\le (1 - \tfrac{\|g\|_2^4}{120\|g\|_1^4}) \|g\|_2^2. \end{aligned}$$

□

By the claim,

$$\begin{aligned} &\gamma^2 \|f_1\|_2^2 + \gamma(1-\gamma)\|\widehat{f}_1.\kappa\|_2^2 \\ &\quad \le \gamma^2 \|f_1\|_2^2 + \gamma(1-\gamma)(1 - \tfrac{\|f_1\|_2^4}{120\|f_1\|_1^4} + \tfrac{10}{N^2})\|f_1\|_2^2 \\ &\quad = \gamma \|\mu_1\|_r^r \big(1 - (1-\gamma)\tfrac{\|\mu_1\|_r^{2r}}{120\|\mu_1\|_{r/2}^{2r}} + \tfrac{10(1-\gamma)}{N^2}\big). \end{aligned}$$

A similar bound holds for f_{-1}. Thus,

$$\begin{aligned} (1) \le \gamma\|\mu_1\|_r^r + (1-\gamma)\|\mu_{-1}\|_r^r - \tfrac{\gamma(1-\gamma)}{120}\big(\tfrac{\|\mu_1\|_r^{3r}}{\|\mu_1\|_{r/2}^{2r}} + \tfrac{\|\mu_{-1}\|_r^{3r}}{\|\mu_{-1}\|_{r/2}^{2r}}\big) \\ + \tfrac{10}{N^2}(\gamma\|\mu_1\|_r^r + (1-\gamma)\|\mu_{-1}\|_r^r). \end{aligned}$$

□

3. Macroscopically

We now analyze the global entropy. We use a *decision tree* to represent the system (Section 3.1). This representation allows to identify positions where entropy is expected to grow (Section 3.2). We then show that all but a tiny fraction of positions indeed increase entropy (Section 3.3).

3.1. **Representing the system.** Think of the elements of $\mathcal{X}$ as vectors $(x_1, x_2, \ldots, x_n)$. Consider a full binary tree of depth n. The root v_0 has depth n and is labeled by the variable x_n. The two children of the root have depth $n-1$ and are labelled by x_{n-1}. In general, all vertices of depth d are labelled by x_d. The depth of the leaves is 0. Every $x \in \mathcal{X}$ defines a walk from the root v_0 to a leaf in the tree. We identify between $\mathcal{X}$ and the leaves in the tree.

Let $\mathcal{B}$ be a collection of vectors in $\mathcal{X}$ of size $|\mathcal{B}| = 2^{\beta n}$. Let P be the uniform distribution on $\mathcal{B}$. The elements of $\mathcal{B}$ correspond to leaves in the tree. Every vertex v in the tree corresponds to the set $\mathcal{B}(v) \subseteq \mathcal{B}$ of all leaves in $\mathcal{B}$ that are under v. Let P_v be the uniform distribution on

$\mathcal{B}(v)$. Let γ_v be the distribution on $\{\pm 1\}$ that is inducted by P_v on the bit from v to its children. If the depth of v is $d = d(v)$ then γ_v is the marginal of P_v on the d'th coordinate.

Fix a parameter $\gamma_0 \in (0, 1/2)$. Call a vertex v *unbiased* if

$$\gamma_v(1) \in [\gamma_0, 1 - \gamma_0].$$

Intuitively, unbiased vertices are positions where the entropy of the system can potentially grow. The unbiased count $\#\mathsf{ub}(v)$ of a vertex v is the number of unbiased vertices on the path from v to the root.

The following claim shows that if $\mathcal{B}$ is large then there are many unbiased vertices.[2]

Claim 8. *For every $\alpha > 0$, the number of leaves v with $\#\mathsf{ub}(v) < \alpha n$ is at most $2^{n(\alpha + H(\alpha) + H(\beta/\log(1/\gamma_0)))}$.*

Proof. Encode a leaf v with $\#\mathsf{ub}(v) < \alpha n$ using the following data:

(1) The depths at which the unbiased nodes appear. There are at most $2^{nH(\alpha)}$ such options.
(2) The value of the path that reaches v at these depths. There are at most $2^{\alpha n}$ such options.
(3) The depths at which the path that reaches v goes through the minority side of a vertex that is not unbiased. If there are δn such depths then $\gamma_0^{\delta n} \geq P(v) = 2^{-\beta n}$. There are, therefore, at most $2^{nH(\beta/\log(1/\gamma_0))}$ such options.

□

3.2. **Where does the entropy grow?** Here we identify positions in the system where the entropy grows. We analyze the entropy of the system for a *fixed* $x \in \mathcal{X}$. In the next section, we see what happens if x is random. We also fix $r \geq 2$ and focus on the ℓ_r norm.

For a vertex v, define a distribution $\mu(v) = \mu_x(v)$ over the integers. If v has depth $d = d(v) > 0$, define $\mu(v)$ to be the distribution of the inner product $\langle B_{\leq d}, x_{\leq d} \rangle$ where $B \sim P_v$. The distribution $\mu(v)$ when v is a leaf gives mass 1 to the integer 0.

Call a vertex v with two children v_1 and v_{-1} *mixing* if it is unbiased and

$$\begin{aligned} \|\mu(v)\|_r^r \leq (1 + \tfrac{10}{N^2})(\gamma_v(1)\|\mu(v_1)\|_r^r + \gamma_v(-1)\|\mu(v_{-1})\|_r^r) \\ - \frac{\gamma_v(1)\gamma_v(-1)}{120}\Big(\frac{\|\mu(v_1)\|_r^{3r}}{\|\mu(v_1)\|_{r/2}^{2r}} + \frac{\|\mu(v_{-1})\|_r^{3r}}{\|\mu(v_{-1})\|_{r/2}^{2r}}\Big). \end{aligned} \tag{4}$$

[2] $H(\xi) = -\xi \log(\xi) - (1-\xi)\log(1-\xi)$ is the binary entropy function.

Intuitively, mixing vertices are places where the entropy strictly increases. This definition makes sense with Lemma 6 in mind.

For a vertex v and a leaf $u \in \mathcal{B}(v)$, denote by $\#\mathsf{mix}(v \to u)$ the number of mixing vertices (including v) on the path from v to u. Let $\alpha_0 > 0$ be a parameter. Recall that v_0 is the root of the tree. Define the set of "good" leaves as

$$\mathcal{G} = \{u \in \mathcal{B} : \#\mathsf{mix}(v_0 \to u) \geq \tfrac{\alpha_0}{4} n\}.$$

Define

$$q(v) = \min\{\#\mathsf{mix}(v \to u) : u \in \mathcal{B}(v) \cap \mathcal{G}\};$$

when $\mathcal{B}(v) \cap \mathcal{G} = \emptyset$, define $q(v) = \min \emptyset = \infty$. A crucial observation is that if $\mathcal{G} \neq \emptyset$ then $q(v_0) \geq \frac{\alpha_0}{4} n$.

The measure $q(v)$ allows to control the entropy of the system at the vertex v. If $q(v)$ is large then the entropy of the system at v is high, as long as $P_v(\neg\mathcal{G})$.

Lemma 9. *Assume $N > 10$. For every $\delta \in (0,1)$, there is a constant $C = C(\gamma_0, \delta) > 0$ so that the following holds. For every $r \geq 2$, every $x \in \mathcal{X}$ and every vertex v so that $\mathcal{B}(v) \neq \emptyset$,*

$$\|\mu(v)\|_r^r \leq (1 + \tfrac{10}{N^2})^{d(v)} \Big(\tfrac{(Cr)^{r/2}}{(q(v) + (Cr)^{r/(r-1)})^{(r-1)/2}} + (\tfrac{1}{1-\delta})^{d(v)} P_v(\neg\mathcal{G}) \Big).$$

The lemma is most interesting at the root v_0:

$$\|\mu(v_0)\|_r^r \leq 3 \Big(\tfrac{(Cr)^{r/2}}{(\alpha_0 n/4)^{(r-1)/2}} + (\tfrac{1}{1-\delta})^n P(\neg\mathcal{G}) \Big). \tag{5}$$

Proof. The proof is by induction. The induction base is when v is a leaf in $\mathcal{B}$. If v is not in $\mathcal{G}$ then $P_v(\neg\mathcal{G}) = 1$ (in this case $q(v) = \infty$). If v is in $\mathcal{G}$ then $q(v) = 0$ and $\frac{(Cr)^{r/2}}{(q(v)+(Cr)^{r/(r-1)})^{(r-1)/2}} = 1$. In both cases, the lemma holds since $\|\mu(v)\|_r^r \leq 1$.

For the induction step, denote by v_1 and v_{-1} the two children of v. Simplify notation:

$$\mu = \mu(v), \quad \mu_1 = \mu(v_1) \quad \& \quad \mu_{-1} = \mu(v_{-1}),$$

and

$$q = q(v), \quad d = d(v) \quad \& \quad \gamma = \gamma_v(1).$$

Since

$$\langle B_{\leq d}, x_{\leq d} \rangle = \begin{cases} \langle B_{<d}, x_{<d} \rangle + x_d & B_d = 1 \\ \langle B_{<d}, x_{<d} \rangle - x_d & B_d = -1 \end{cases}$$

we can write

$$\mu = \gamma S_{x_d} \mu_1 + (1-\gamma) S_{-x_d} \mu_{-1}.$$

There are two cases to consider.

Non-mixing. If v is not mixing then $q = \min\{q_1, q_{-1}\}$. Since $P_v(\neg\mathcal{G}) = \gamma P_{v_1}(\neg\mathcal{G}) + (1-\gamma)P_{v_{-1}}(\neg\mathcal{G})$, Observation 5 and induction imply

$$\begin{aligned}\|\mu\|_r^r &\leq \gamma\|\mu_1\|_r^r + (1-\gamma)\|\mu_{-1}\|_r^r \\ &\leq (1+\tfrac{10}{N^2})^d\big(\tfrac{(Cr)^{r/2}}{(q+(Cr)^{r/(r-1)})^{(r-1)/2}} + (\tfrac{1}{1-\delta})^d P_v(\neg\mathcal{G})\big).\end{aligned}$$

Mixing. If v is mixing then $q = 1 + \min\{q_1, q_{-1}\}$. The following claim summarizes the main technical part.

Claim 10. *The following bound holds for v_1:*

$$(1+\tfrac{10}{N^2})\|\mu_1\|_r^r - \tfrac{1-\gamma}{120}\tfrac{\|\mu_1\|_r^{3r}}{\|\mu_1\|_{r/2}^{2r}} \tag{6}$$

$$\leq (1+\tfrac{10}{N^2})^d\big(\tfrac{(Cr)^{r/2}}{(q+(Cr)^{r/(r-1)})^{(r-1)/2}} + (\tfrac{1}{1-\delta})^d P_{v_1}(\neg\mathcal{G})\big). \tag{7}$$

A similar bound holds for v_{-1}.

Proof. For simplicity of notation, let

$$\eta = q - 1 + (Cr)^{r/(r-1)} \quad \& \quad C' = \tfrac{9}{\delta^2}(Cr)^{r(r-2)/(r-1)}.$$

Start by considering the case that

$$(\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G}) > \tfrac{1-\delta}{\delta}\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}}.$$

In this case,

$$\begin{aligned}\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}} + (\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G}) &< P_{v_1}(\neg\mathcal{G})(\tfrac{1}{1-\delta})^{d-1}(\tfrac{\delta}{1-\delta}+1) \\ &= P_{v_1}(\neg\mathcal{G})(\tfrac{1}{1-\delta})^d.\end{aligned}$$

By induction,

$$\|\mu_1\|_r^r \leq (1+\tfrac{10}{N^2})^{d-1}\big(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}} + (\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G})\big).$$

This completes the proof:

$$\begin{aligned}(6) &\leq (1+\tfrac{10}{N^2})\|\mu_1\|_r^r \\ &\leq (1+\tfrac{10}{N^2})^d\big(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}} + (\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G})\big) \\ &\leq (1+\tfrac{10}{N^2})^d P_{v_1}(\neg\mathcal{G})(\tfrac{1}{1-\delta})^d \leq (7).\end{aligned}$$

We can thus assume that

$$(\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G}) \leq \tfrac{1-\delta}{\delta}\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}}.$$

Convexity implies

$$\|\mu_1\|_r^r = \sum_z \mu_1(z)\big((\mu_1(z))^{\frac{r-2}{2}}\big)^{\frac{2(r-1)}{r-2}} \geq (\|\mu_1\|_{r/2}^{r/2})^{\frac{2(r-1)}{r-2}}.$$

By induction and the above,

$$\begin{aligned}\|\mu_1\|_{r/2}^{2r} &\le \left(\|\mu_1\|_r^r\right)^{\frac{2(r-2)}{r-1}} \\ &\le \left((1+\tfrac{10}{N^2})^{d-1}\left(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}} + (\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G})\right)\right)^{\frac{2(r-2)}{r-1}} \\ &\le \tfrac{C'}{\eta^{r-2}}.\end{aligned}$$

Hence,

$$\tfrac{1-\gamma}{120}\tfrac{\|\mu_1\|_r^{3r}}{\|\mu_1\|_{r/2}^{2r}} \ge \tfrac{1-\gamma}{120C'}\eta^{r-2}\|\mu_1\|_r^{3r}.$$

So,

$$(6) \le (1+\tfrac{10}{N^2})\|\mu_1\|_r^r - \tfrac{1-\gamma}{120C'}\eta^{r-2}\|\mu_1\|_r^{3r}.$$

The map $\xi \mapsto (1+\frac{10}{N^2})\xi - \frac{1-\gamma}{120C'}\eta^{r-2}\xi^3$ is increasing when $\xi^2 \le \frac{C'}{\eta^{r-2}}$. Since $\eta \ge (Cr)^{r/(r-1)}$,

$$\begin{aligned}\left((1+\tfrac{10}{N^2})^{d-1}\left(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}} + (\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G})\right)\right)^2 &\le \left(\tfrac{3}{\delta}\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}}\right)^2 \\ &\le \tfrac{9}{\delta^2}\tfrac{(Cr)^r}{\eta^{r-2}}\cdot\tfrac{1}{\eta} \\ &\le \tfrac{C'}{\eta^{r-2}}.\end{aligned}$$

Therefore, by induction

$$\begin{aligned}(6) &\le (1+\tfrac{10}{N^2})^d\left(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}} + (\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G})\right) \\ &\quad - \tfrac{1-\gamma}{120C'}\eta^{r-2}\left((1+\tfrac{10}{N^2})^{d-1}\left(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}} + (\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G})\right)\right)^3 \\ &\le (1+\tfrac{10}{N^2})^d(\tfrac{1}{1-\delta})^{d-1}P_{v_1}(\neg\mathcal{G}) \\ &\quad + (1+\tfrac{10}{N^2})^d\left(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}} - \tfrac{1-\gamma}{360C'}\eta^{r-2}\left(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}}\right)^3\right).\end{aligned}$$

Focus on the expression inside the last brackets:

$$\begin{aligned}&\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}}\left(1 - \tfrac{(1-\gamma)\delta^2(Cr)^{r(r-2)/(r-1)}}{360\cdot 9}\eta^{r-2}\left(\tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}}\right)^2\right) \\ &\quad = \tfrac{(Cr)^{r/2}}{\eta^{(r-1)/2}}\left(1 - \tfrac{(1-\gamma)\delta^2(Cr)^{r/(r-1)}}{3240}\tfrac{1}{\eta}\right).\end{aligned}$$

Since v is unbiased, to complete the proof, it suffices to show that

$$\tfrac{1}{\eta^{(r-1)/2}}\left(1 - \tfrac{\gamma_0\delta^2(Cr)^{r/(r-1)}}{3240}\tfrac{1}{\eta}\right) \le \tfrac{1}{(\eta+1)^{(r-1)/2}}$$

or

$$\left(1+\tfrac{1}{\eta}\right)^{(r-1)/2}\left(1 - \tfrac{\gamma_0\delta^2(Cr)^{r/(r-1)}}{3240}\tfrac{1}{\eta}\right) \le 1.$$

Since $\eta \geq 2r$ and since $0 \leq \xi \leq \frac{1}{2k}$ implies $(1+\xi)^k \leq 1 + 2k\xi$,

$$\Big(1+\tfrac{1}{\eta}\Big)^{(r-1)/2}\Big(1-\tfrac{\gamma_0\delta^2(Cr)^{r/(r-1)}}{3240}\tfrac{1}{\eta}\Big) \leq \Big(1+\tfrac{2r}{\eta}\Big)\Big(1-\tfrac{\gamma_0\delta^2(Cr)^{r/(r-1)}}{3240}\tfrac{1}{\eta}\Big)$$
$$\leq 1 + \tfrac{2r}{\eta} - \tfrac{\gamma_0\delta^2(Cr)^{r/(r-1)}}{3240}\tfrac{1}{\eta} \leq 1.$$

□

The claim completes the proof of the lemma, since

$$\begin{aligned}\|\mu\|_r^r &\leq (1+\tfrac{10}{N^2})(\gamma\|\mu_1\|_r^r + (1-\gamma)\|\mu_{-1}\|_r^r) \\ &\quad - \tfrac{\gamma(1-\gamma)}{120}\Big(\tfrac{\|\mu_1\|_r^{3r}}{\|\mu_1\|_{r/2}^{2r}} + \tfrac{\|\mu_{-1}\|_r^{3r}}{\|\mu_{-1}\|_{r/2}^{2r}}\Big) \\ &= \gamma\Big((1+\tfrac{10}{N^2})\|\mu_1\|_r^r - \tfrac{1-\gamma}{120}\tfrac{\|\mu_1\|_r^{3r}}{\|\mu_1\|_{r/2}^{2r}}\Big) \\ &\quad + (1-\gamma)\Big((1+\tfrac{10}{N^2})\|\mu_{-1}\|_r^r - \tfrac{\gamma}{120}\tfrac{\|\mu_{-1}\|_r^{3r}}{\|\mu_{-1}\|_{r/2}^{2r}}\Big).\end{aligned}$$

□

3.3. **Many mixing leaves.** The previous section (Lemma 9) highlights the role of the good leaves $\mathcal{G}$ in the overall entropy of the system. To show that the overall entropy is high, we need to show that $\mathcal{G}$ is typically almost full.

Lemma 11. *There is a constant $c = c(\alpha_0) > 0$ so that*

$$\mathop{\mathbb{E}}_X[P(\neg\mathcal{G})] < 2^{-cn} + \frac{2^{n(\alpha_0 + H(\alpha_0) + H(\beta/\log(1/\gamma_0)))}}{|\mathcal{B}|}.$$

Proof.

$$\mathop{\mathbb{E}}_X[P(\neg\mathcal{G})] = \frac{1}{|\mathcal{B}|}\sum_{v\in\mathcal{B}}\mathop{\mathbb{P}}_X[\#\mathsf{mix}(v) < \tfrac{\alpha_0}{4}n].$$

By Claim 8, the number of leaves v with $\#\mathsf{ub}(v) < \alpha_0 n$ is at most $2^{n(\alpha_0+H(\alpha_0)+H(\beta/\log(1/\gamma_0)))}$. We can thus focus on the rest of the leaves. Let v be a leaf in $\mathcal{B}$ with $K = \#\mathsf{ub}(v) \geq \alpha_0 n$. Let $u_1, \ldots, u_K$ be the unbiased vertices on the path from the root to v. Denote by E_k the indicator random variable for the event that u_k is mixing (namely, (4) holds for u_k).

Claim 12. *For each $k > 1$, we have $\mathbb{E}[E_k | E_1, \ldots, E_{k-1}] \geq \frac{1}{2}$.*

Proof. Fix $u = u_k$ of depth d. Denote its two children by u_1 and u_{-1}. Fix $X_{<d}$ so that $\mu_X(u_1)$ and $\mu_X(u_{-1})$ are fixed as well. Let X_d be uniform in $\{\pm 1\}$. Lemma 6 implies that for at least one choice of X_d the vertex u is mixing. □

By the claim, the sequence of random variables S_0 and $S_k = E_1 + E_2 + \ldots + E_k - \frac{k}{2}$ is a submartingale. Azuma's inequality implies that

$$\mathbb{P}[\#\mathsf{mix}(v) < \tfrac{\alpha_0}{4} n] \leq \mathbb{P}[S_K - S_0 < -\tfrac{K}{4}] \leq e^{-\frac{K}{32}}.$$

□

3.4. **Putting it together.**

Proof of Theorem 3. Let X be uniformly distributed in $\mathcal{X}$. Let $\mathcal{B}$ be a family of vectors in $\mathcal{X}$ of size $|\mathcal{B}| = 2^{\beta n}$. Let $\alpha_0, \gamma_0 > 0$ be so that $4(\alpha_0 + H(\alpha_0)) = \beta$ and $4H(\beta/\log(1/\gamma_0)) = \beta$. By Lemma 11, there is a constant $c' = c'(\beta) > 0$ so that

$$\mathop{\mathbb{E}}_X[P(\neg\mathcal{G})] < 2^{-c'n} + \tfrac{2^{n(\alpha_0 + H(\alpha_0) + H(\beta/\log(1/\gamma_0)))}}{|\mathcal{B}|} \leq 2^{-3cn}$$

for some $c = c(\beta) > 0$. By Markov's inequality

$$\mathop{\mathbb{P}}_X[P(\neg\mathcal{G}) > 2^{-2cn}] < 2^{-cn}.$$

Choose $\delta = \delta(\beta) > 0$ so that $(\frac{1}{1-\delta})^n = 2^{cn}$. By (5), there is a constant $C' = C'(\beta) > 0$ so that

$$\|\mu_X\|_r^r \leq 3\big(\tfrac{(C'r)^{r/2}}{(\alpha_0 n/4)^{(r-1)/2}} + 2^{cn} P(\neg\mathcal{G})\big).$$

It follows that

$$\mathbb{P}[\|\mu_X\|_r^r > 3\big(\tfrac{(C'r)^{r/2}}{(\alpha_0 n/4)^{(r-1)/2}} + 2^{-cn}\big)] < 2^{-cn}.$$

There is $C = C(\beta) > 0$ so that

$$3\big(\tfrac{(C'r)^{r/2}}{(\alpha_0 n/4)^{(r-1)/2}} + 2^{-cn}\big) \leq C^r \tfrac{r^{r/2}}{n^{(r-1)/2}}.$$

□

Proof of Theorem 1. By Theorem 3, there are $c, C > 0$ so that

$$\mathop{\mathbb{P}}_X[\|\mu_X\|_2^2 > \tfrac{C}{\sqrt{n}}] < C2^{-cn}.$$

Since $\mathbb{P}[X \in \mathcal{A}] \geq 2^{-cn/2}$,

$$\mathop{\mathbb{P}}_A\left[\|\mu_A\|_2^2 > \tfrac{C}{\sqrt{n}}\right] \leq C2^{-cn}2^{cn/2} = C2^{-cn/2}.$$

By Claim 2,

$$\mathop{\mathbb{P}}_A\left[\mu_A(I) > \sqrt{|I|\tfrac{C}{\sqrt{n}}}\right] \leq \mathop{\mathbb{P}}_A[\|\mu_A\|_2^2 > \tfrac{C}{\sqrt{n}}].$$

Hence,

$$\mathop{\mathbb{P}}_{A,B}[\langle A, B\rangle \in I] = \mathop{\mathbb{P}}_A[\mu_A(I)] \leq \sqrt{|I|\tfrac{C}{\sqrt{n}}} + C2^{-cn/2}.$$

□

Acknowledgement. I wish to thank Anup Rao, Oded Regev, Avishay Tal and David Woodruff for helpful conversations.

Appendix A. Norms and anti-concentration

Proof of Claim 2.

$$\|\nu\|_r^r \geq |I| \sum_{\omega \in I} \tfrac{1}{|I|}(\nu(\omega))^r \geq |I| \Big(\sum_{\omega \in I} \tfrac{1}{|I|}\nu(\omega) \Big)^r = \tfrac{1}{|I|^{r-1}}(\nu(I))^r.$$

□

References

[1] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. In *STOC*, pages 333–342, 2011.

[2] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *FOCS*, pages 337–347, 1986.

[3] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

[4] A. Chakrabarti and O. Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM Journal on Computing*, 41(5):1299–1317, 2012.

[5] E. Chattopadhyay and D. Zuckerman. Explicit two-source extractors and resilient functions. In *STOC*, pages 670–683, 2016.

[6] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[7] P. Erdös. On a lemma of littlewood and offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.

[8] P. Frankl and Z. Furedi. Solution of the littlewood-offord problem in high dimensions. *Annals of Mathematics*, pages 259–270, 1988.

[9] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica*, 8(3-4):197–211, 1977.

[10] P. Indyk and D. Woodruff. Tight lower bounds for the distinct elements problem. In *FOCS*, pages 283–288, 2003.

[11] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation (iii). *Rec. Math. (Mat. Sbornik) N.S*, 12(3):277–286, 1943.

[12] R. Raz. Extractors with weak random seeds. In *STOC*, pages 11–20, 2005.

[13] A. A. Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(1):197–208, 2012.

[14] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81(1):73–205, 1995.

[15] T. Tao and V. Vu. A sharp inverse littlewood-offord theorem. *Random Structures & Algorithms*, 37(4):525–539, 2010.

[16] T. Tao and V. H. Vu. Inverse littlewood-offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, pages 595–632, 2009.
[17] L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *FOCS*, pages 32–42, 2000.
[18] T. Vidick. A concentration inequality for the overlap of a vector on a large set. *Chicago Journal of Theoretical Computer Science*, 1:1–12, 2012.

Department of Mathematics, Technion-IIT
Email address: amir.yehudayoff@gmail.com

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il