Lecture 13: Hypercontractivity and Small-Set Expansion

January 31, 2023

Lecturer: Max Hopkins *Scribe: Max Hopkins*

In this lecture, we overview the theory of *hypercontractivity* and its applications to understanding expansion of higher order random walks beyond Cheeger's constant and the second eigenvalue.

1 Review: Basic Fourier Analysis on HDX

Last lecture, we discussed an elegant theory of Fourier analysis on high dimensional expanders, and in particular the existence of a basis for functions $f : X(k) \to \mathbb{R}$ on k-faces[1] of an HDX

$$f = \sum_{i=0}^{k} f_i,$$

where f_i is the 'contribution' to f coming from i-faces, or more explicitly $f_i = U_i^k g_i$ for some $g_i \in Ker(D_i)$. We showed that this basis satisfies the following useful properties:

1. **Approximate Eigenbasis**: $Mf_i \approx_\gamma \lambda_i f_i$
2. **Approximate Orthogonality**: $\forall i \neq j : \langle f_i, f_j \rangle \approx_\gamma 0$
3. **Approximate Parseval**: $\langle f, f \rangle \approx_\gamma \sum_{i=0}^{k} \langle f_i, f_i \rangle$

We discussed briefly how this matches the classical theory of Fourier analysis on the cube, even up to the particular values of λ_i for the lower walk!

$$\lambda_i \approx \frac{k-i}{k}.$$

The fact that this exactly matches the (lazy) hypercube graph is not a coincidence, and we will see a general connection later in this lecture.

Finally, recall we discussed briefly one of the reasons this theory is useful. If we can understand how a function projects onto this decomposition, we understand its expansion. In particular, for a set $S \subseteq X(k)$, write $f = 1_S$ and observe:

$$\bar{\Phi}(S) = \frac{\langle f, Mf \rangle}{\langle f, f \rangle} \approx \sum_{i=0}^{k} \lambda_i \frac{\langle f, f_i \rangle}{\langle f, f \rangle}.$$

[1] *N.b.* in this lecture, we use $X(k)$ to denote k-sets, *not* $(k+1)$-*sets*. We will see this is the correct formulation when one wishes to generalize Fourier analysis from the cube.

We already understand the values of λ_i, which are determined by the inherent laziness of the underlying poset architecture. In this lecture, we will see how to understand the level-i Fourier weight:

$$W_i(f) = \frac{\langle f, f_i \rangle}{\langle f, f \rangle},$$

for the special case of simplicial complexes, completing this direction.

2 On The Structure of Non-Expanding Sets

With this in mind, let's take a step back. We are interested in understanding the expansion of functions on HDX (with respect to say the lower walk), but what does this really mean? One way to formalize this into an interesting question is through its contrapositive:

If $\bar{\Phi}(S) \geqslant \delta$, what can we infer about the structure of S?

One reason this formulation is interesting is that it often reduces to natural property testing questions that can be useful e.g. for PCP reductions. To be concrete, let's consider the case of the standard hypercube graph:

Definition 2.1 (Hypercube Graph)**.** The n-dimensional hypercube graph has vertex set $V = \{0,1\}^n$, and edges between vertices of hamming distance 1. Equivalently, we can think of the random walk on $V = \{0,1\}^n$ that flips the value of a uniformly random coordinate.

The hypercube has a 'canonical' family of sparse cuts (non-expanding sets). Recall that a function $f : \{0,1\}^n \to \mathbb{R}$ is called an *i-junta* if its value only depends on some subset of i coordinates. The expansion of any i-junta on the hypercube graph is at most:

$$\Phi(i\text{-junta}) \leqslant \frac{i}{n},$$

since this is the probability we flip a coordinate in the sensitive subset. The property testing question is now the converse: is *every* sparse cut a junta? This type of result turns out to be very useful when designing PCP reductions that use (for instance) the hypercube as an inner gadget.

In fact, before it was used for PCPs and hardness of approximation, this formulation of our question was perhaps *the* foundational question in the modern theory of boolean function analysis. Here, non-expanding functions on the cube were referred to as functions with 'low total influence,' and their structure was famously described in seminal works of Kahn, Kalai, and Linial [KKL88], and Friedgut [Fri98].

Theorem 2.2 (Friedgut's Junta Theorem [Fri98])**.** *If $\Phi(S) \leqslant \frac{i}{n}$, then S is ε-close to a $2^{O(i/\varepsilon)}$-junta.*

Today, we will focus on a simpler setting that also generalizes more directly to product spaces and high dimensional expanders,[2] small-set expansion and the *noisy cube*.[3]

[2] Friedgut's Junta Theorem is false on products. There is a generalization called Hatami's Theorem to objects called 'pseudojuntas', but even the definition is fairly involved. Extending this result to HDX is an open problem.

[3] We take slight liberty with the definition here so that it matches the noise operator. Often the noisy cube is defined by flipping each bit with probability ρ.

Definition 2.3 (Noisy Hypercube)**.** The n-dimensional ρ-noisy hypercube is the graph on vertex set $V = \{0,1\}^n$ whose edge distribution around $x \in V$ is described by the following process:

1. Remove each bit from x with probability $1 - \rho$.

2. Re-sample each removed bit uniformly at random.

We call the random-walk operator of this graph the **Noise Operator**, and denote it by T_ρ.

We'll use $\Phi_\rho(S)$ to denote expansion with respect to the T_ρ. Ahlswede and Gács [AG76], and later Kahn, Kalai, and Linial [KKL88] observed that the noisy hypercube is a *small-set expander*, giving essentially the strongest possible characterization of non-expanding sets in this regime.

Theorem 2.4 (Small-Set Expansion Theorem)**.** *The expansion of any set S of density α with respect to the noisy hypercube T_ρ is:*[4]

$$\Phi_\rho(S) \geqslant 1 - \alpha^{\frac{1-\rho}{1+\rho}}.$$

In other words, every non-expanding set is large:

$$\bar{\Phi}_\rho(S) \geqslant \delta \implies \mathbb{E}[1_S] \geqslant \delta^{O(1)}$$

Small-set expansion is a very special type of behavior on graphs, with deep connections to topics such as locally testable codes, hardness of approximation, and the unique games conjecture. We know very few classes of objects satisfying (even variants of) this type of characterization, and building a general theory is a major open problem.

2.1 Beyond the Cube

That said, we do have at least a few analogs of the small-set expansion theorem beyond the cube. Related characterizations on the p-biased cube, shortcode, and Grassmannian, for instance, have lead to numerous breakthroughs in hardness of approximation and the theory of sharp thresholds, including the recent proof of the 2-2 Games Conjecture [KMS18]. Today, we'll focus on the simplest of these extensions, the *p-biased cube* and *product spaces*, along with their relation to HDX.

Recall the p-biased cube is simply the distribution over $\{0,1\}^n$ given by drawing each bit independently from Ber(p).[5] We can define natural analogs of the noise operator on the p-biased cube (and indeed general product spaces), where the only change is that the re-sampling step is performed with respect to the apporpriate distribution instead of uniformly at random (so in this case, each bit will be re-sampled from Ber(p)).

It turns out that the *spectrum* of analogous graphs on p-biased cube (indeed on any n-dimensional product) are identical to the standard cube. So naively, one might hope this object exhibits the same properties beyond spectral analysis. Unfortunately, it only takes a moments thought to refute this conjecture.

[4]Notice that the exponent here perfectly interpolates between the $\rho = 0$ setting, where we re-sample the entire string and therefore hit S with probability α, and the $\rho = 1$ setting where we don't re-sample at all and therefore always stay in S.

[5]I.e. each bit is 1 independently with probability p.

Example 2.5 (p-biased cube)**.** *Consider a dictator function on the p-biased cube:*

$$\mathbf{1}_i = \{x \in \{0,1\}^n : \ x_i = 1\}$$

The density of any dictator is p. On the other hand:

$$\Phi_\rho(\mathbf{1}_i) \leqslant 1 - \rho,$$

since we can only leave the dictator if we re-sample coordinate i.

In other words, the noisy p-biased cube is *not* a small-set expander, and once again **local functions** such as dictators and juntas provide examples of (now small) non-expanding sets. Naively, one might expect some variant of Friedgut's Junta Theorem holds, but it turns out this is too much to hope for.[6] We can, however, prove a weaker result that's still quite useful: every non-expanding function is indeed *local* in the sense that it must have constant correlation with a low-dimensional restriction.

More formally, we will actually state this result for general product spaces. Let $X = (\Omega^{\otimes i}, \pi^{\otimes i})$ be an n-dimensional product. We call a function $f : X \to \mathbb{R}$ ε-*global* if it is small under every low-dimensional conditioning. That is if for all coordinate sets $S \subseteq [n]$, $|S| = \frac{1}{2}\log(\varepsilon^{-1})$ and labelings $z_S \in \Omega^S$:

$$\|f_{S \to z_S}\|_2^2 := \mathbb{E}[f(x)^2 \mid x_S = z_S] \leqslant \varepsilon,$$

In the language of complexes (which we'll discuss next section), this simply means that f is sparse over all low-dimensional links.

Theorem 2.6 (Global sets expand [KLLM19])**.** *Every ε-global function $f : X \to \{0,1\}$ expands near perfectly:*

$$\Phi_\rho(f) \geqslant 1 - \varepsilon^{O(1)}.$$

*In other words, any non-expanding set is **local**:*

$$\Phi_\rho(f) \leqslant \delta \implies \exists S, z : \|f_{S \to z}\|_2^2 \geqslant \delta^{O(1)}$$

where $|S| = O(\log(1/\delta))$.

This weaker type of characterization still turns out to be quite useful, and can be used to show sharp thresholds,[7] as well as various further results in boolean function analysis and extremal combinatorics [LM19, KLLM21]. A final interesting observation is that while this result may appear to be weaker than the small-set expansion theorem at first glance, it is actually a direct generalization![8] This is because on the cube, *every* sparse function is global.

Observation 2.7. *Every function $f : \{0,1\}^n \to \{0,1\}$ of density α is $\alpha^{O(1)}$-global.*

We leave the proof as a simple exercise.[9]

[6] Consider OR for small enough p.

[7] Here technically a stronger variant is required which holds for monotone functions on the p-biased cube, see [KLLM19]

[8] At least if one is willing to be loose in the exact exponent.

[9] Here's another interesting exercise: use this idea to recover hypercontractivity on HDX that scales with the worst marginal probability of any vertex (across all restrictions). This type of result is called a generalized hypercontractivity Theorem, and was also used on the p-biased cube to derive early results on sharp thresholds.

2.2 High Dimensional Expanders

Product spaces have a very natural interpretation as partite simplicial complexes. In particular, given an n-dimensional product space $(\Omega, \pi) = (\Omega_1 \times \ldots \times \Omega_n, \pi_1 \times \ldots \times \pi_n)$, we can think of (Ω, π) as an n-dimensional partite simplicial complex on vertex set

$$X(1) = \bigcup_{i=1}^{n} \{i\} \times \Omega_i$$

where coordinates in the original space correspond to the parts or 'colors' of the partite complex, and top-dimensional faces/weights simply correspond to their analogous string in the product. Indeed, this is really a general translation between any joint distribution on n variables, and an n-partite simplicial complex.

Viewed in this fashion, products have very natural connections to the theory we've developed in this class. In fact, perhaps the first connection to observe is that products are 'perfect' HDX.

Observation 2.8 (Products are HDX)**.** *Products are (one-sided) 0-local-spectral expanders.*

Proof. If each (Ω_i, π_i) is uniform, all links are unweighted complete multi-partite graphs, which are (one-sided) 0-spectral expanders. We leave the case of general π_i as an exercise. □

In fact, the connection runs much deeper. Many standard analysis tools on products are also specific examples of the theory of higher order random walks. Consider, for simplicity, the setting of the hypercube. The (non-lazy) down-up walks on its corresponding complex generate the Hamming scheme (e.g. standard hypercube graph). In this vein, there is also a natural generalization of the noise operator in this language:

$$T_\rho = \sum_{i=0}^{n} \binom{n}{i} (1-\rho)^i \rho^{n-i} U_{n-i}^{n} D_{n-i}^{n}.$$

This operator can be read as performing the following procedure, starting at an n-face x:

1. Remove each coordinate with probability $1 - \rho$, denote set of removed coordinates S
2. Re-sample x_S conditioned on $x_{[n]\setminus S}$

It is not hard to check that this is exactly the noise operator when applied to a product. It can even be checked that the spectrum of these operators remains (roughly) the same as on a product, and depends only on the underlying dimension!

This raises a natural question. Since products and the noise operator are a special class of walks on HDX,

Can we hope to generalize the theory of small-set expansion to HDX?

It certainly isn't obvious HDX will exhibit the same type of structure (given say the differences between the uniform and p-biased cube). Moreover bounded-degree complexes actually act quite differently with respect to classically related properties such as fast mixing.[10] Despite all this, it turns out a direct generalization is indeed possible!

[10]Mixing, small-set expansion, and hypercontractivity are intimately related topics, see e.g. [Mon04]. Products mix in time dependent only on dimension, while bounded-degree HDX suffer a log factor in the number of vertices due to degree considerations.

Theorem 2.9 ([GLL21, HKL23]). *Let $X(n)$ be a partite γ-HDX,*[11] *then any ε-global function expands near perfectly:*

$$\Phi_\rho(f) \geqslant 1 - \varepsilon^{O(1)}.$$

In other words, any non-expanding set is ***local****:*

$$\Phi_\rho(f) \leqslant \delta \implies \exists S, z : \|f_{S \to z}\|_2 \geqslant \delta^{O(1)}$$

where $|S| = O(\log(1/\delta))$.

This covers bounded-degree high dimensional expanders such as the Kaufman-Oppenheim construction we saw earlier in this class. We note that a similar statement also holds in the general (non-partite) setting for *two-sided* local spectral expanders [BHKL21]. However, it is possible to recover this result from the above just by embedding an unordered complex into a partite one,[12] so we focus just on the partite setting.

It is an intriguing open question whether Theorem 2.9 has further applications in areas such as hardness of approximation, sharp thresholds, and graph theory where traditional (and extended) tools in boolean function analysis drove many years of breakthroughs. In the $\mathbb{F}_2$-regime, a similar type of structure has recently seen use in derandomizing classical hardness results for Sum-of-Squares [DFHT20, HL22], but no direct applications of Theorem 2.9 itself are known.

2.2.1 HDX are 'Product-like'

Given the substantial differences between products and bounded degree complexes, how can we hope to prove Theorem 2.9? Is there some sense in which HDX act like products up to an error term? To answer this question, it will be convenient to think of a partite complex $X(n)$ as a *joint distribution* over n variables (given by the values on each color/coordinate). The condition that $X(n)$ is a product is exactly when these variables are *independent*. On the other hand, this is extremely far from true on any bounded-degree complex. If we condition on a coordinate taking some value, there are only a constant number of values the other coordinates can take!

The key lies in observing that while this is true on a point-wise level, the variables actually are close to independent *spectrally*.[13] More formally, given disjoint coordinate sets $S, T \subseteq [n]$, consider the induced marginal distributions over X_S and X_T (equivalently the complexes induced by restricting to color sets S and T). The correlation between X_S and X_T is captured by the following averaging operator $E_{S,T} : C(X_T, \mathbb{R}) \to C(X_S, \mathbb{R})$ mapping functions on X_T to functions on X_S by averaging over neighbors in the complex:

$$E_{S,T} f(x_S) = \mathop{\mathbb{E}}_z[f(z_T) | z_S = x_S].$$

Equivalently we can view $E_{S,T}$ as the random walk moving between X_S and X_T via shared faces in $X_{S \cup T}$. On a product, $E_{S,T}$ is a complete bipartite graph since there are no correlations between X_S and X_T. On an HDX, Dikstein and Dinur [DD19] proved that $E_{S,T}$ is an excellent expander!

[11] Here and throughout, we will always mean one-sided local-spectral expanders by this notation.

[12] In particular, include every permutation of each top-dimensional face in the original complex.

[13] This is also very closely related to the popular notion of 'spectral independence' on spin-systems [ALO20, CGSV21], the only difference is that the spin-system version looks at the aggregate operator across all colors.

Theorem 2.10 (Theorem 7.1 [DD19])**.** *If X is an n-dimensional partite γ-HDX, then:*[14]

$$\lambda(E_{S,T}) \leqslant O(|S||T|\gamma).$$

Proof. We leave the proof as an exercise.[15] □

In other words, while individual settings of S and T may be extremely dependent, on average it is actually possible to decorrelate them. We will see later in this lecture how this simple observation can be applied to generalize the theory of Fourier analysis on products to HDX, a result due to Tom Gur, Noam Lifshitz, and Siqi Liu [GLL21].[16]

3 Hypercontractivity on the Cube

We are finally ready to discuss the powerful toolset used to prove these small-set expansion theorems: *hypercontractivity*. The basic theory of Fourier analysis we've covered so far, e.g. orthogonality and Parseval, largely pertain to *second moment* or *spectral* properties of the underlying complex. Hypercontractivity is a tool used to control the behavior of functions *beyond* the second moment. We will use a simple, but central form of hypercontractivity called the Bonami Lemma:

Theorem 3.1 (The Bonami Lemma)**.** *For any $f : \{0,1\}^n \to \mathbb{R}$, and any $i \leqslant n$:*

$$\|f^{\leqslant i}\|_4 \leqslant \sqrt{3}^i \|f^{\leqslant i}\|_2,$$

where we recall $f^{\leqslant i} = \sum_{|S| \leqslant i} \widehat{f}(S)\chi_S$ is the degree at most i part of f.

It is worth taking a moment to interpret the above result. In the boolean setting, $f^{\leqslant i}$ can be an arbitrary multilinear polynomial of degree i. Thus the Bonami lemma can be interpreted as saying:

"Low-degree polynomials are smooth,"

since 'spiky' functions get blown up by higher norms. Another useful interpretation is as a niceness condition on random variables (that the 4th moment doesn't exceed the 2nd too drastically). Indeed, one can use the Bonami Lemma to derive very useful probabilistic results such as tail bounds (extensions of Chernoff-Hoeffding), and even anti-concentration (see [O'D14, Chapter 9]).

Why is this useful in our case? It's not too hard to observe that sparse functions are *not* smooth. A boolean function of density α always satisfies

$$\|f\|_4 = \alpha^{1/4} \gg \alpha^{1/2} = \|f\|_2.$$

The Bonami lemma, intuitively, should then imply that a sparse function cannot have much projection onto its low degree Fourier coefficients! This can be made formal by a clever application of Holder's inequality.

[14] Here $\lambda(E_{S,T})$ stands for the second largest singular value of $E_{S,T}$, like in HW 1.
[15] Hint: prove a partite version of Trickling-Down, and use the fact that $\lambda(E_{S,T}) = \lambda(E_{T,S})$.
[16] Formally GLL introduced this notion separately as an object called a 'γ-product,' and observed that embedded *two-sided* HDX satisfy the definition.

Lemma 3.2 (Level-i inequality). *Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function of density α, then:*

$$\frac{\langle f, f_i\rangle}{\langle f, f\rangle} \leqslant 3^i \alpha^{1/2}$$

Proof.

$$\begin{aligned}
\langle f, f_i\rangle &\leqslant \|f_i\|_4 \|f\|_{4/3} && \text{(Holder's Inequality)}\\
&\leqslant \sqrt{3}^i \|f_i\|_2 \|f\|_{4/3} && \text{(Bonami Lemma)}\\
&= \sqrt{3}^i \sqrt{\langle f, f_i\rangle} \|f\|_{4/3} && \text{(Orthogonality)}\\
&= \sqrt{3}^i \sqrt{\langle f, f_i\rangle} \alpha^{3/4} && \text{(Booleanity)}
\end{aligned}$$

Re-arranging then gives

$$\langle f, f_i\rangle \leqslant 3^i \alpha^{3/2},$$

and observing that $\alpha = \langle f, f\rangle$ by Booleanity gives the desired result. □

It's worth noting that Holder's Inequality is frequently an indispensible tool when working with hypercontractivity and higher norms in general. We leave it as an exercise to deduce the small-set expansion theorem given this lemma.[17]

Why 'hypercontractivity'? As a brief aside before we prove the Bonami lemma, one might wonder why this result is called hypercontractivity. In fact, this really refers to an equivalent formulation of the lemma that concerns the contraction of the noise operator T_ρ:

Theorem 3.3 (The $(2,4)$-Hypercontractivity Theorem). *For any $\rho \leqslant \frac{1}{\sqrt{3}}$ and $f : \{0,1\}^n \to \mathbb{R}$:*

$$\|T_\rho f\|_4 \leqslant \|f\|_2$$

This result can be interpreted as saying not only does T_ρ contract ℓ_2-norm (just by dint of being an averaging operator), it contracts *higher* norms into ℓ_2 (hence the term 'hyper'). It is not particularly hard to deduce Theorem 3.3 from the Bonami Lemma (or prove directly by induction as below). We refer the reader to [O'D14] and leave both as exercises.

3.1 Proving the Bonami Lemma

We now give a simple proof of the Bonami lemma, which also serves as an excellent example of how many results in boolean function analysis can be proved via induction on the dimension n. Since we are in the boolean setting it is equivalent to show that for any degree-d multilinear polynomial $g : \{-1,1\}^n \to \mathbb{R}$:

$$\mathbb{E}[g^4] \leqslant 9^d \mathbb{E}[g^2]^2.$$

The key is to observe that by splitting g into terms by their dependence on the final variable x_n, we can write:

$$g = x_n g' + g''$$

where g' is degree $d-1$, g'' is degree d, and both are functions on only $n-1$ variables. This facilitates induction on n.

[17] Here we really mean a weak version giving $\Phi(S) \geqslant 1 - \alpha^{O(1)}$, reaching the tight exponent is more involved.

Base Case: The base case is $n = 0$, which are simply constant functions. Here the result holds trivially since $\mathbb{E}[c^4] = \mathbb{E}[c^2]^2 = c^4$.

Inductive Step: Let's first examine the 2-norm:

$$\mathbb{E}[g^2] = \mathbb{E}[x_n^2 g'^2] + \mathbb{E}[2x_n g' g''] + \mathbb{E}[g''^2] = \mathbb{E}[g'^2] + \mathbb{E}[g''^2]$$

since $x_n^2 = 1$, $\mathbb{E}[x_n] = 0$, and g' is independent of x_n. We can expand the 4-norm similarly, setting $x_n^2 = 1$ and killing terms with odd dependence on x_n:

$$\begin{aligned}\mathbb{E}[g^4] &= \mathbb{E}[g'^4] + 6\mathbb{E}[g'^2 g''^2] + \mathbb{E}[g''^4] \\ &\leqslant \mathbb{E}[g'^4] + 6\sqrt{\mathbb{E}[g'^4]\mathbb{E}[g''^4]} + \mathbb{E}[g''^4]\end{aligned}$$

We can now apply the inductive hypothesis to get

$$\begin{aligned}\mathbb{E}[g^4] &\leqslant \mathbb{E}[g'^4] + 6\sqrt{\mathbb{E}[g'^4]\mathbb{E}[g''^4]} + \mathbb{E}[g''^4] \\ &\leqslant 9^{k-1}\mathbb{E}[g'^2]^2 + 6\sqrt{9^{k-1}\mathbb{E}[g'^2]^2 9^k \mathbb{E}[g''^2]^2} + 9^k\mathbb{E}[g''^2]^2 \\ &\leqslant 9^k(\mathbb{E}[g'^2]^2 + 2\mathbb{E}[g'^2](\mathbb{E}[g''^2] + \mathbb{E}[g''^2]^2) \\ &= 9^k\mathbb{E}[g^2]^2\end{aligned}$$

4 Hypercontractivity Beyond the Cube

In this section we introduce Fourier analysis on products and discuss its extension to HDX.

4.1 Hypercontractivity and The Efron-Stein Decomposition

The Bonami Lemma only has meaning if we have a notion of a decomposition of f into components by 'degree.' This role will be filled by a generalization of the typical Fourier basis on the cube to products called the *Efron-Stein Decompisition*. For an extended exposition, see [O'D14, Chapter 8].

Like the classical Fourier decomposition, Efron-Stein breaks f into contributions coming from each subset of coordinates:

$$f = \sum_{S \subseteq [n]} f^{=S},$$

where on the cube $f^{=S}$ is $\widehat{f}(S)\chi_S$. In this setting, however, we will take a *combinatorial* approach to defining each contribution $f^{=S}$. Towards this end, it will be useful to define a set of *colored averaging operators*, which will play a similar role to the up and down operator in the linear algebraic decomposition we saw last lecture. Given any subset $S \subseteq [n]$, define E_S to be the operator that averages a function f over its values on S:

$$E_S[f](x) := \mathbb{E}[f(z) \mid z_S = x_S].$$

Equivalently, E_S can be thought of as re-randomizing f over all coordinates outside of S. E_S satisfies many useful properties. One can check for instance...

Claim 4.1 (Some Properties of E_S)**.** *On any partite complex X:*

1. *E_S contracts p-norms*
2. *E_S is self-adjoint*

In a sense E_S can be viewed as a restricted version of the down-up walk (where the down-process is forced to walk to coordinate set S). In fact, this can be made formal by the following observation:

$$U_i^n D_i^n = \frac{1}{\binom{n}{i}} \sum_{|S|=i} E_S.$$

Back to the task at hand, what is the contribution to f coming from S? One natural idea is just to use $E_S f$. This is almost right, but the expression inherently counts contributions coming from all subsets of S as well. Similar to the linear algebraic approach, we want to subtract out such contributions. Using inclusion-exclusion, this suggests the following formula:

$$f^{=S} = \sum_{T \subseteq S} (-1)^{|S \setminus T|} E_T f$$

We leave it to the reader to verify that this definition does indeed lead to a decomposition of f.

Theorem 4.2 (Efron-Stein Decomposition)**.** *For any n-dimensional product space X and function $f : X \to \mathbb{R}$, the Efron-Stein decomposition satisfies:*

1. ***Eigenbasis**: $T_\rho f^{=S} = \rho^{|S|} f^{=S}$*
2. ***Orthogonality**: $\forall S \neq S' : \langle f^{=S}, f^{=S'} \rangle = 0$*
3. ***Parseval**: $\langle f, f \rangle = \sum\limits_{i=0}^{n} \langle f^{=S}, f^{=S} \rangle$*

The proof of these results relies on the following simple but key observation:

Claim 4.3. *For any n-dimensional product space X and any subsets $S, T \subseteq [n]$:*

$$E_S E_T = E_{S \cap T}$$

Proof. Recall E_T re-randomizes a function over $[n] \setminus T$. Applying E_S and E_T, we are re-randomizing over all coordinates except $S \cap T$. Since X is a product these operations are independent, so this is exactly $E_{S \cap T}$. □

In fact, this is the *only* assumption we will make on X (besides being a partite complex), which hints at what's to come!

Proof of Theorem 4.2. Parseval is immediate from orthogonality.

Eiegenbasis: We can write T_ρ in terms of the averaging operators as:

$$T_\rho = \sum_{T \subseteq [n]} (1-\rho)^{n-|T|} \rho^{|T|} E_T$$

Then we can write:

$$T_\rho f^{=S} = \sum_{T \subseteq [n]} (1-\rho)^{n-|T|} \rho^{|T|} E_T f^{=S}$$

Let's examine $E_T f^{=S}$ separately, expanding out $f^{=S}$ we get

$$\begin{aligned}
E_T f^{=S} &= E_T \left(\sum_{T' \subseteq S} (-1)^{|S \setminus T'|} E_{T'}[f] \right) \\
&= \left(\sum_{T' \subseteq S} (-1)^{|S \setminus T'|} E_T E_{T'}[f] \right) && \text{(Linearity of } E_T\text{)} \\
&= \left(\sum_{T' \subseteq S} (-1)^{|S \setminus T'|} E_{T \cap T'}[f] \right) && \text{(Claim 4.3)}
\end{aligned}$$

Note that when $T \not\supseteq S$, this sum is clearly 0.[18] Putting both together, we have:

$$T_\rho f^{=S} = \sum_{T \supseteq S} (1-\rho)^{n-|T|} \rho^{|T|} \left(\sum_{T' \subseteq S} (-1)^{|S \setminus T'|} E_{T \cap T'}[f] \right)$$

Since $T \supseteq S$ and $T' \subseteq S$, we always have $T \cap T' = T'$ so the righthand sum is just $f^{=S}$. Therefore

$$T_\rho f^{=S} = \left(\sum_{T \supseteq S} (1-\rho)^{n-|T|} \rho^{|T|} \right) \cdot f^{=S},$$

The lefthand sum is just the probability that a ρ-biased string contains S, which is exactly $\rho^{|S|}$.

Orthogonality: We now wish to show that for any $S \neq S'$:

$$\langle f^{=S}, f^{=S'} \rangle = 0$$

This follows from the self-adjointness of the operator E_T. Assume without loss of generality that there exists some $x \in S \setminus S'$, then we can write:

$$\begin{aligned}
\langle f^{=S}, f^{=S'} \rangle &= \sum_{T \subseteq S \setminus \{x\}} (-1)^{|S \setminus T|} \langle (E_T - E_{T \cup \{x\}}) f, f^{=S'} \rangle \\
&= \sum_{T \subseteq S \setminus \{x\}} (-1)^{|S \setminus T|} \langle f, (E_T - E_{T \cup \{x\}}) f^{=S'} \rangle.
\end{aligned}$$

But since $x \notin S'$ by definition

$$\begin{aligned}
(E_T - E_{T \cup \{x\}}) f^{=S'} &= \sum_{T' \subseteq S'} (-1)^{|S \setminus T'|} (E_T - E_{T \cup \{x\}}) E'_T f \\
&= \sum_{T' \subseteq S'} (-1)^{|S \setminus T'|} (E_{T \cap T'} - E_{(T \cup \{x\}) \cap T'}) f && \text{(Claim 4.3)} \\
&= 0 && (x \notin T')
\end{aligned}$$

□

[18]Why? To see this, try fixing an element in $x \in S \setminus T$ and splitting the sum into terms $T' \subseteq S \setminus \{x\}$ and $T' \cup \{x\}$.

The Efron-Stein Decomposition has many nice properties beyond those stated here, such as its behavior under restriction. As above, these typically depend only on the partite structure of X, and (perhaps repeated applications of) Claim 4.3. Later in the lecture, we will see how to use these facts to prove hypercontractivity for global functions.

Theorem 4.4 (Bonami Lemma for Products [KLLM19, Zha21])**.** *Any $f \in L^2(\Omega^n, \pi^{\otimes n})$ satisfies:*

$$\|f^{\leqslant i}\|_4^4 \leqslant 2^{O(i)} \|f\|_2^2 \max_{|S|=i, z_S} \{\|f_{S \to z}\|_2^2\}.$$

Expansion of global functions follows similarly as on the noisy cube. We note that Theorem 4.4 is also sometimes called *conditional* hypercontractivity, due to its dependence on the sparsity of f over X's conditional distributions.

As a brief historical aside, we note Theorem 4.4 was actually proved independently by Keevash, Lifshitz, Long, and Minzer [KLLM19], and O'Donnell and Zhao [Zha21]. The latter only appeared in Zhao's thesis,[19] so is lesser known, but their approach generalizes very nicely to HDX.

4.1.1 From Products to HDX

The Efron-Stein decomposition generalizes very naturally to arbitrary partite simplicial complexes. In this section, we will show how to use the 'spectral' independence of HDX to derive approximate variants of the same properties. At a high level, the following is a good guiding principle:

Any classical property of Efron-Stein holds approximately on HDX.

More concretely, following [GLL21], we extend the central claim in the prior section to HDX:

Claim 4.5 (Lemma 3.3 [GLL21])**.** *For any n-dimensional partite γ-HDX X and any subsets $S, T \subseteq [n]$:*

$$\|E_S E_T - E_{S \cap T}\|_2 \leqslant O(|S||T|\gamma)$$

Proof. We first argue we can reduce to the case where $S \cap T = \emptyset$, simply by moving to the link of $T \cap T'$. To see this, first write:

$$\mathbb{E}_x[(E_S E_T f - E_{S \cap T} f)^2] = \mathbb{E}_{x_{S \cap T}} \mathbb{E}_{x_{[n] \setminus (S \cap T)}}[(E_S E_T f(x_{S \cap T}, \cdot) - E_{S \cap T} f(x_{S \cap T}, \cdot))^2]. \quad (1)$$

We now compare these restrictions to the averaging operators on the link of $x_{S \cap T}$ itself, which we denote by $E_Y^{x_{X \cap T}}$ for $Y \subset [n] \setminus S \cap T$. The latter restriction is simple, it is just the expectation of f over the link of $x_{S \cap T}$ by definition, so

$$E_{S \cap T} f(x_{S \cap T}, \cdot) = E_{\emptyset}^{x_{T \cap S}}(f_{x_{S \cap T}}).$$

The former restriction is slightly more complicated, we wish to argue

$$E_S E_T f(x_{S \cap T}, \cdot) = E_{S \setminus T}^{x_{S \cap T}} E_{T \setminus S}^{x_{S \cap T}}(f_{x_{S \cap T}})$$

[19]While this was a few years later, the result and proof technique were actually mentioned in Zhao's thesis proposal contemporaneously with [KLLM19].

Intuitively, this is just because in both E_S and E_T, I am averaging within the link of $x_{S\cap T}$. Since the conditional distributions are defined appropriately, I can just perform this averaging inside the link of $x_{S\cap T}$ instead. Formally, this is a pain to write out, but it is a good exercise in playing with conditional/link probabilities.

Returning to (1), we now have:

$$\begin{aligned}\mathbb{E}_x[(E_S E_T f - E_{S\cap T} f)^2] &= \mathbb{E}_{v=x_{S\cap T}}\mathbb{E}_{x_{[n]\setminus(S\cap T)}}[(E^v_{S\setminus T}E^v_{T\setminus S}f_v - E^v_{\varnothing}f_v)^2]\\ &\leqslant |S|^2|T|^2\gamma^2\mathbb{E}_v[\|f_v\|_2^2]\\ &= |S|^2|T|^2\gamma^2\|f\|_2^2,\end{aligned}$$

assuming the result holds for the non-intersecting setting in all links. As such, it is enough to prove for any non-intersecting sets S,T:

$$\|E_S E_T - E_{\varnothing}\|_2 \leqslant |T||T'|\gamma.$$

The trick is now to notice that $E_T E_{T'} = E_{T',T}E_{T'}$ and $\mathbb{E}[E_{T'}f] = \mathbb{E}[f]$, so

$$\begin{aligned}\|E_T E_{T'} f - E_{\varnothing} f\|_2 &= \|E_{T',T}(E_{T'}f) - E_{\varnothing}(E_{T'}f)\|_2\\ &\leqslant |T'||T|\gamma\|E_{T'}f\|_2\\ &\leqslant |T'||T|\gamma\|f\|_2\end{aligned}$$

since averaging contracts 2-norms. □

As an immediate corollary, we get a version of Theorem 4.2 for HDX.

Corollary 4.6 (Efron-Stein on HDX [GLL21]). *For any n-dimensional partite γ-HDX X and function $f : X \to \mathbb{R}$, the Efron-Stein decomposition satisfies:*

1. ***Approximate Eigenbasis**: $T_\rho f^{=S} \approx_\gamma \rho^{|S|} f^{=S}$*
2. ***Approximate Orthogonality**: $\forall S \neq S' : \langle f^{=S}, f^{=S'}\rangle \approx_\gamma 0$*
3. ***Approximate Parseval**: $\langle f, f\rangle \approx_\gamma \sum\limits_{i=0}^{n}\langle f^{=S}, f^{=S}\rangle$*

Proof. Repeat exactly the same arguments as in Theorem 4.2, but replace every use of Theorem 4.3 with Theorem 4.5 at the cost of $O(\gamma)$ error. □

With this in mind, we can now state a natural variant of the Bonami Lemma for HDX

Theorem 4.7 (Bonami Lemma for HDX (Informal [HKL23])). *Let f be any function on an n-dimensional partite γ-HDX. Then as long as $\gamma \ll 2^{-poly(n)}$:*

$$\|f^{\leqslant i}\|_4^4 \lesssim_\gamma 2^{O(i)}\|f^{\leqslant i}\|_2^2 \max_{|S|=i, z_S}\{\|f_{S\to z}\|_2^2\}$$

While this exact variant of the result does not appear in the literature, one can see [BHKL21, GLL21] for formal proofs of very similar results.

5 Hypercontractivity (The Proof!)

We are finally ready to sketch the proof of the hypercontractivity theorem. We will focus mostly on the basic setting of product spaces, but discuss throughout how to generalize the proof to HDX.

5.1 Symmetrization (Reduction to the Cube)

While the Efron-Stein basis is very useful in its own right, and can be analyzed directly, it still tends to be much harder to deal with than the typical Fourier expansion over $\{0, 1\}^n$. Largely for this reason, the main technique used in the study of products are methods to reduce analysis back to the cube. In this section, we overview an elegant framework for this technique called *symmetrization* due to Kahane [Kah68] and Bourgain [Bou79]. Our exposition largely follows [O'D14, Chapter 10].

Definition 5.1. Let $f \in L^2(\Omega^n, \pi^{\otimes n})$ be any function over a product space. The symmetrization of f, $\tilde{f} \in L^2(\{-1, 1\}^n \times \Omega^n, \pi_{1/2}^{\otimes n} \times \pi^{\otimes n})$ is:

$$\tilde{f}(r, x) = \sum_{S \subseteq [n]} r_S f^{=S}(x)$$

where $\{f^{=S}\}_{S \subset [n]}$ is the standard orthogonal decomposition and $r_S = \prod_{j \in S} r_j$

Broadly speaking, the idea behind this definition lies in the hope that hitting each component with a random binary string won't significantly change the distribution of f. Indeed, when f is the cube this intuition can be made formal, and it's not hard to see that f and $\tilde{f}$ are equi-distributed. While this is no longer true on products, it is true that the moments of f and $\tilde{f}$ remain closely related. Let's start with the second moment, where we can actually show that f and $\tilde{f}$ are exactly equivalent.

Proposition 5.2. *Let $f \in L^2(\Omega^n, \pi^{\otimes n})$ be any function over a product space. Then f and $\tilde{f}$ have the same second moment:*

$$\|f\|_2 = ||\tilde{f}||_2$$

Proof. The trick is to notice that the restriction of $\tilde{f}$ to any $x \in \Omega^n$ is a boolean function with Fourier coefficients:

$$\widehat{\tilde{f}|_x}(S) = f^{=S}(x).$$

By Parseval's Theorem we therefore have for all $x \in \Omega^n$:

$$\|\tilde{f}|_x\|_2^2 = \sum_{S \subset [n]} f^{=S}(x)^2$$

Taking the expectation on both sides over $x \sim \pi^{\otimes n}$ then gives the desired result by orthogonality of $f^{=S}$. On an HDX we can use approximate Parseval instead. □

This is a useful result in and of itself, but we are mostly interested in analyzing higher moments. Bourgain's key observation in this regime is that while higher moments are not necessarily equivalent, they are bounded on both sides by an application of the noise operator

Theorem 5.3 (The Symmetrization Theorem [Bou79]). *Let $f \in L^2(\Omega^n, \pi^{\otimes n})$ be any function over a product space and $q > 1$. Then the qth moment of f is sandwiched by symmetrized applications of the noise operator:*

$$\|\widetilde{T_{c_q} f}\|_q \leqslant \|f\|_q \leqslant \|\widetilde{T_2 f}\|_q$$

for some constant $0 \leqslant c_q \leqslant 1$ dependent only on q.

For our purposes, it will actually be sufficient to show just the upper bound, which is still somewhat non-trivial. We will assume the result in the univariate case (which can be used as is for the HDX setting as well), and prove the result by induction.

Lemma 5.4 (Univariate Symmetrization). *For any single-variate $h \in L^2(\Omega, \pi)$:*

$$\|T_{1/2} h(x)\|_q \leqslant \|\tilde{h}(x)\|_q, \tag{2}$$

Before moving to the full proof, it will be useful to introduce some notation, namely a coordinate wise variant of the noise operator T^i_ρ, which re-samples the ith coordinate of f with probability $1 - \rho$. On a product, this can be equivalently be written as:

$$T^i_\rho f = \sum_{S \not\ni i} f^{=S} + \sum_{S \ni i} \rho f^{=S}.$$

A similar result can be proved on HDX by using spectral independence of coordinates.

It will further be useful to extend this definition beyond the domain $\rho \in [0, 1]$, which can be done in the latter formulation. Then the coordinate-wise 'noise' operator also carries a close connection with symmetrization, namely for $r = (r_1, \ldots, r_n) \in \{-1, 1\}^n$:

$$T_r f(x) = T^1_{r_1} \ldots T^n_{r_n} f(x) = \sum_{S \subset [n]} r_S f^{=S}(x) = \tilde{f}(r, x).$$

Proof of Theorem 5.3 (Upper Bound). First we argue the univariate result can be extended to show the result holds for the single coordinate operator:

$$\|T^i_{1/2} f(x)\|_q \leqslant \|T^i_{r_i} f\|_q.$$

Assume for simplicity $i = 1$, and write $x = (x_1, x')$. The idea is to fix x' and analyze $f|_{x'}$ which is a function on a single variable:

$$\|T^i_{1/2} f(x)\|_q = \|\|(T^i_{1/2} f)|_{x'}(x_1)\|_{q,x_1}\|_{q,x'} = \|\|T_{1/2} f|_{x'}(x_1)\|_{q,x_1}\|_{q,x'}.$$

Here the final equality relies on the fact that the coordinate-wise noise operator respects restriction to that coordinate. This is a good exercise to show for products, and is again possible to extend approximately to HDX. We can now apply Equation (2):

$$\begin{aligned} \|\|T_{1/2} f|_{x'}(x_1)\|_{q,x_1}\|_{q,x'} &\leqslant \|\|T_{r_i} f|_{x'}(x_1)\|_{q,x_1}\|_{q,x'} \\ &= \|T^i_{r_i} f(x)\|_q \end{aligned}$$

where we have again used that $T^i_{r_i}$ respects restriction. Since we have proved this coordinate-wise result for general functions, the full result follows from a basic induction. The idea is simply to apply the coordinate-wise version iteratively to the outmost coordinate (e.g. to analyze $T^1_{1/2} T^2_{1/2} f$, you'd apply the coordinate version to $T^1_{1/2}$ on function $T^2_{1/2} f$), noting that we can apply this to all coordinates since the coordinate-wise operators commute and may therefore be re-ordered at will. On an HDX, the operators only approximately commute, but this is still enough for an approximate version of the result to go through. ☐

5.2 Hypercontractivity

We are now ready to prove hypercontractivity for products/HDX. We follow the proof of O'Donnell and Zhao [O'D14].

Theorem 5.5 (Bonami Lemma for Products). *Let $f \in L^2(\Omega^n, \pi^{\otimes n})$ be an (ε, i)-pseudorandom function. Then the fourth moment of $f^{\leqslant i}$ is upper bounded by:*

$$\|f^{\leqslant i}\|_4^4 \leqslant 2^{O(i)}\|f\|_2^2 \max_{|S|=i, z_S}\{\|f_{S\to z}\|_2^2\}.$$

Proof. The main idea is to use Bourgain's symmetrization trick combined with the standard Bonami lemma for the discrete hypercube. In particular, writing $g = T_2 f^{\leqslant i}$ for notational simplicity, notice that we can write:

$$\begin{aligned}
\mathbb{E}_x[(f^{\leqslant i})^4] &\leqslant \mathbb{E}_x\left[\mathbb{E}_r[\tilde{g}|_x(r)^4]\right] \\
&\leqslant 2^{O(i)}\mathbb{E}_x\left[\mathbb{E}_r[\tilde{g}(x,r)^2]^2\right] && \text{(Bonami Lemma)} \\
&= 2^{O(i)}\mathbb{E}_x\left[\left(\sum_{|S|\leqslant i} g^{=S}(x)^2\right)^2\right] && \text{(Parseval)}
\end{aligned}$$

where the final step follows from recalling that $g^{=S}(x)$ are the Fourier coefficients of $\tilde{g}(x,\cdot)$. Recalling $g = T_2 f^{\leqslant i}$, we therefore have:

$$\mathbb{E}_x[(f^{\leqslant i})^4] \leqslant 2^{O(i)}\mathbb{E}_x\left[\left(\sum_{|S|\leqslant i} 2^{2|S|} f^{=S}(x)^2\right)^2\right] \leqslant 2^{O(i)}\mathbb{E}_x\left[\left(\sum_{|S|\leqslant i} f^{=S}(x)^2\right)^2\right]$$

The analysis from this point is essentially a much simpler version of [KMMS18]: we'll expand out the above sum over the intersection of index sets $I = S \cap T$, pull out one of the resulting terms by its maximum (which is then bounded by pseudorandomness), and note that the remaining term is simply the 2-norm. Let's start by re-indexing our sum over the intersection I and pulling these variables outside the summation:

$$\begin{aligned}
\mathbb{E}_x\left[\left(\sum_{|S|\leqslant i} f^{=S}(x)^2\right)^2\right] &= \mathbb{E}_x\left[\sum_{|I|\leqslant i}\sum_{S\supset I:|S|\leqslant i} f^{=S}(x)^2\left(\sum_{T:|T|\leqslant i, S\cap T=I} f^{=T}(x)^2\right)\right] \\
&\leqslant \sum_{|I|\leqslant i}\mathbb{E}_{x_I}\left[\left(\sum_{S\supset I:|S|\leqslant i}\mathbb{E}_{x_{S\setminus I}}\left[f^{=S}(x_S)^2\right]\right)\left(\sum_{T\supset I:|T|\leqslant i}\mathbb{E}_{x_{T\setminus I}}\left[f^{=T}(x_T)^2\right]\right)\right]
\end{aligned}$$

where we have additionally used product structure to push the relevant variables inside each sum. We use x_S and x_T since $f^{=S}$ and $f^{=T}$ depend only on these variables. The idea is now to pull out the second term by bounding its maximum over I:

$$\mathbb{E}\left[\left(\sum_{|S|\leqslant i} f^{=S}(x)^2\right)^2\right] \leqslant \sum_{|I|\leqslant i} \mathbb{E}_{x_I}\left[\left(\sum_{S\supset I:|S|\leqslant i} \mathbb{E}_{x_{S\setminus I}}\left[f^{=S}(x_S)^2\right]\right)\right] \max_{|I|\leqslant i, y_I\in\Omega_I}\left(\sum_{T\supset I:|T|\leqslant i} \mathbb{E}_{x_{T\setminus I}}\left[f^{=T}(y_I, x_{T\setminus I})^2\right]\right)$$

$$= \left(\sum_{|I|\leqslant i}\sum_{S\supset I:|S|\leqslant i} \mathbb{E}_{x_S}\left[f^{=S}(x_S)^2\right]\right) \max_{|I|\leqslant i, y_I\in\Omega_I}\left(\sum_{T\supset I:|T|\leqslant i} \mathbb{E}_{x_{T\setminus I}}\left[f^{=T}(y_I, x_{T\setminus I})^2\right]\right)$$

$$\leqslant 2^i \|f^{\leqslant i}\|_2^2 \max_{|I|\leqslant i, y_I\in\Omega_I}\left(\sum_{T\supset I:|T|\leqslant i} \mathbb{E}_{x_{T\setminus I}}\left[f^{=T}(y_I, x_{T\setminus I})^2\right]\right)$$

where we have observed that each $f^{=S}$ term in the first summation appears exactly $2^{|S|} \leqslant 2^i$ times (once for each subset of S). It is left to bound the maximum term, which follows from the following relation on restrictions of the orthogonal decomposition:

$$f^{=I\cup B}(y_I, x_B) = \sum_{J\subset I} (-1)^{|I|-|J|} (f_{J|_{y_J}})^{=B}(x_B),$$

where $f_{J|_{y_J}} : \Omega^{[n]\setminus J} \to \mathbb{R}$ is the restriction of f obtained by setting coordinates J to corresponding values y_J. Cauchy-Schwarz then gives:

$$\sum_{T\supset I:|T|\leqslant i} \mathbb{E}_{x_{T\setminus I}}\left[f^{=T}(y_I, x_{T\setminus I})^2\right] \leqslant \sum_{T\supset I} \mathbb{E}_{x_{T\setminus I}}\left[f^{=T}(y_I, x_{T\setminus I})^2\right]$$

$$= \sum_{T\supset I} \mathbb{E}_{x_{T\setminus I}}\left[\left(\sum_{J\subset I} (-1)^{|I|-|J|} (f_{J|_{y_J}})^{=T\setminus I}(x_{T\setminus I})\right)^2\right]$$

$$\leqslant 2^i \sum_{J\subset I}\sum_{S\subset \bar{I}} \mathbb{E}_{x_S}[(f_{J|_{y_J}})^{=S}(x_S)^2]$$

$$\leqslant 4^i \varepsilon \|f\|_\infty^2$$

where the last step follows from noting that each of the at most 2^i restricted sums is upper bounded by the restricted two-norm $\|f_{J|_{y_J}}\|_2^2 \leqslant \varepsilon\|f\|_\infty^2$ by the pseudorandomness of f. Plugging this back into our previous computations gives

$$\mathbb{E}[(f^{\leqslant i})^4] \leqslant 2^{O(i)} \varepsilon \|f\|_2^2 \|f\|_\infty^2$$

as desired. □

References

[AG76] Rudolf Ahlswede and Peter Gács. Spreading of sets in product spaces and hypercontraction of the markov operator. *The annals of probability*, pages 925–939, 1976.

[ALO20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. *arXiv preprint arXiv:2001.00303*, 2020.

[BHKL21] Mitali Bafna, Max Hopkins, Tali Kaufman, and Shachar Lovett. Hypercontractivity on high dimensional expanders. *arXiv preprint arXiv:2111.09444*, 2021.

[Bou79] J Bourgain. Walsh subspaces of lp-product spaces. *Séminaire danalyse fonctionnelle de lécole Polytechnique. Palaiseau*, 1979.

[CGSV21] Zongchen Chen, Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1548–1557. SIAM, 2021.

[DD19] Yotam Dikstein and Irit Dinur. Agreement testing theorems on layered set systems. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1495–1524. IEEE, 2019.

[DFHT20] Irit Dinur, Yuval Filmus, Prahladh Harsha, and Madhur Tulsiani. Explicit sos lower bounds from high-dimensional expanders. *arXiv preprint arXiv:2009.05218*, 2020.

[Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[GLL21] Tom Gur, Noam Lifshitz, and Siqi Liu. Hypercontractivity on high dimensional expanders. *arXiv preprint arXiv:2111.09375*, 2021.

[HKL23] Max Hopkins, Tali Kaufman, and Shachar Lovett. Symmetrization and higher moments on high dimensional expanders. 2023.

[HL22] Max Hopkins and Ting-Chun Lin. Explicit lower bounds against $\Omega(n)$-rounds of sum-of-squares. *arXiv preprint arXiv:2204.11469*, 2022.

[Kah68] JP Kahane. Some random series of functions, dc heath and co. *Raytheon Education Co., Lexington, Mass*, 1968.

[KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 68–80, Los Alamitos, CA, USA, oct 1988. IEEE Computer Society.

[KLLM19] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Hypercontractivity for global functions and sharp thresholds. *arXiv preprint arXiv:1906.05568*, 2019.

[KLLM21] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Global hypercontractivity and its applications. *arXiv preprint arXiv:2103.04604*, 2021.

[KMMS18] Subhash Khot, Dor Minzer, Dana Moshkovitz, and Muli Safra. Small set expansion in the johnson graph. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 25, page 78, 2018.

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. Pseudorandom sets in grassmann graph have near-perfect expansion. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 592–601. IEEE, 2018.

[LM19] Noam Lifshitz and Dor Minzer. Noise sensitivity on the p-biased hypercube. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1205–1226. IEEE, 2019.

[Mon04] Ravi Montenegro. Convergence of markov chains, September 2004.

[O'D14] Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[Zha21] Yu Zhao. *Generalizations and Applications of Hypercontractivity and Small-Set Expansion*. PhD thesis, Columbia University, 2021.

A Claim 4.5 Details

Since our notation departs somewhat from the rest of the course, we include the proof in Claim 4.5 that:

$$E_S E_T f(x_{S\cap T}, \cdot) = E_{S\setminus T}^{x_{S\cap T}} E_{T\setminus S}^{x_{S\cap T}} (f_{x_{S\cap T}})$$

To prove this formally, it is perhaps easier to start with the restriction and unroll the definitions:

$$E_{S\setminus T}^{v} E_{T\setminus S}^{v} (f_{x_{S\cap T}})(y) = \mathop{\mathbb{E}}_{z|v} \left[E_{T\setminus S}^{v} f_v(z) \mid z_{S\setminus T} = y_{S\setminus T} \right]$$

Now examining the inner term, we can unroll the restriction:

$$\begin{aligned} E_{T\setminus S}^{v} f_v(z) &= \mathop{\mathbb{E}}_{z'|v} [f_v(z') \mid z'_{T\setminus S} = z_{T\setminus S}] \\ &= \mathop{\mathbb{E}}_{z'|v} [f(v, z') \mid z'_{T\setminus S} = z_{T\setminus S}] \\ &= \mathop{\mathbb{E}}_{z''} [f(z'') \mid z''_T = (v, z_{T\setminus S})] \end{aligned}$$

Plugging this back into the above gives

$$\begin{aligned} \mathop{\mathbb{E}}_{z|v} \left[\mathop{\mathbb{E}}_{z''} [f(z'') \mid z''_T = (v, z_{T\setminus S})] \mid z_{S\setminus T} = y_{S\setminus T} \right] &= \mathop{\mathbb{E}}_{z} \left[\mathop{\mathbb{E}}_{z''} [f(z'') \mid z''_T = z_T] \mid z_S = (v, y_{S\setminus T}) \right] \\ &= \mathop{\mathbb{E}}_{z} \left[E_T f \mid z_S = (v, y_{S\setminus T}) \right] \\ &= E_S E_T f(v, \cdot) \end{aligned}$$