Independent Sets in Graph Powers are Almost Contained in Juntas

Irit Dinur * Ehud Friedgut † Oded Regev ‡

January 7, 2008

Abstract

Let $G = (V, E)$ be a simple undirected graph. Define G^n, the n-th power of G, as the graph on the vertex set V^n in which two vertices $(u_1, \dots, u_n)$ and $(v_1, \dots, v_n)$ are adjacent if and only if u_i is adjacent to v_i in G for every i. We give a characterization of all independent sets in such graphs whenever G is connected and non-bipartite.

Consider the stationary measure of the simple random walk on G^n. We show that every independent set is almost contained with respect to this measure in a junta, a cylinder of constant co-dimension. Moreover we show that the projection of that junta defines a nearly independent set, i.e., it spans few edges (this also guarantees that it is not trivially the entire vertex-set).

Our approach is based on an analog of Fourier analysis for product spaces combined with spectral techniques and on a powerful invariance principle presented in [18]. This principle has already been shown in [11] to imply that independent sets in such graph products have an influential coordinate. In this work we prove that in fact there is a set of few coordinates and a junta on them that capture the independent set almost completely.

1 Introduction

The n-th *power* of an undirected graph $G = (V, E)$, denoted by G^n, is defined as follows: the vertex set is V^n and two vertices $(u_1, \dots, u_n)$ and $(v_1, \dots, v_n)$ are adjacent in G^n if and only if u_i is adjacent to v_i in G for every i. This is, in graph-theoretic terms, the n-fold weak product of G with itself. An alternative description, which is helpful for the spectral approach we wish to adopt, is that the adjacency matrix of G^n is the n-fold tensor product of the adjacency matrix of G with itself.

In this paper we study independent sets in G^n where G remains fixed while n tends to infinity. In classical graph theory one studies the size of the maximal independent set in a graph or, adopting analytical language, its measure according to the uniform measure on the set of vertices. In the case of G^n this is also the product measure defined by the uniform measure on G. It turns out that there is an alternative measure on G such that the corresponding product measure on G^n is very

AMS 2000 subject classification: 05d05

Key words and phrases: Independent sets, Intersecting families, Product graphs, Discrete Fourier analysis.

*School of Computer Science and Engineering, Hebrew University, Jerusalem, Israel. email: dinuri at cs.huji.ac.il

†Institute of Mathematics, Hebrew University, Jerusalem, Israel. email: ehudf at math.huji.ac.il. Research supported in part by the Israel Science Foundation, grant no. 0329745.

‡Department of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Supported by an Alon Fellowship, by the Binational Science Foundation, by the Israel Science Foundation, and by the European Commission under the Integrated Project QAP funded by the IST directorate as Contract Number 015848.

well suited to studying independent sets. This is the stationary measure of the simple random walk on G, whence the product measure on G^n is the stationary measure of the simple random walk on G^n.

A *junta* is a set $J \subseteq V^n$ that is described by a constant (independent of n) number of coordinates, i.e., a cylinder of constant co-dimension. For example let $I \subseteq V$ be an independent set in G. The set $I \times V^{n-1}$ of all vertices whose first coordinate belongs to I is a junta. This set is also an independent set in the graph G^n. Another example of a relatively large independent set in G^n is the set of all vectors that have at least two of their first three coordinates in I:

$$S = \{(v_1, v_2, \ldots, v_n) \in V^n \ : \ \text{at least two of } v_1, v_2, v_3 \text{ belong to } I\} \ .$$

If μ is any probability measure on G and $\mu(I) = \alpha$ then it is easy to see that S is an independent set in G^n, whose measure is $3\alpha^2 - 2\alpha^3$ according to the product measure.

Seeing these two examples might lead one to conjecture that the only reasonably large independent sets are juntas. A moment of reflection shows that this conjecture is too naive for several reasons. The first is that, as usual in these settings, one should modify the statement to say that all large independent sets are *approximable* by juntas - one can add small perturbations to independent sets to achieve other examples. The second difficulty, that makes the proof of the statement much trickier, is that any subset of an independent set is also independent. So, for example, a random subset of our previous examples evades complete description by a small set of coordinates. However this example is still *contained* in a junta. A modified conjecture would be, then, that every independent set can be approximated by an independent set that is contained in a junta. Of course for the theorem to be meaningful we need the junta to be non-trivial (i.e., not all of V^n which, trivially, "depends on few coordinates"). We will ensure this by showing that the junta itself, which depends on j coordinates, is close to being an independent set in the corresponding graph on V^j.

Still, as alluded to above, for certain graphs G this principle is not true if the underlying measure according to which we measure our approximation is the uniform measure. Let us study an example of this. Let G be K_4 minus an edge, a graph on vertex set $\{a, b, c, d\}$ where all pairs of vertices except $\{b, c\}$ are edges. Consider the following independent set in G^n:

$$I = \{v \in \{a, b, c, d\}^n \ : \ \text{more than half the coordinates of } v \text{ are in } \{b, c\}\}.$$

Obviously I is an independent set, for any two vertices in it must share a coordinate where their entries do not span an edge in G. The number of vertices in I is close to half of 4^n (depending on the parity of n), i.e., it captures asymptotically half of the vertices in G^n. Also, clearly, I is not close in the uniform measure to any junta. The "reason" that I evades the principle we are aiming to prove is that the average degree of the vertices in I is much lower than that of G^n, hence it is "easy" for I to be independent. (The average degree of the vertices in I is less than $(\sqrt{6})^n$ whereas the average degree in G^n is $(\frac{5}{2})^n$.) Hence a reasonable measure to consider, which might imply more structure on independent sets, is one where the measure of vertices is proportional to their degree - this is, of course, a measure that arises naturally: it is the stationary measure of the simple random walk on G^n. This, finally, turns out to be the correct setting for our main theorem.

Let $G = (V, E)$ be a simple, undirected, connected, non-bipartite graph. Throughout this paper $\mu(\cdot)$ will always denote the unique stationary measure of the simple random walk on G:

$$\mu(u) = \frac{\deg(u)}{2|E|}.$$

By abuse of notation we will often use μ to also denote the product measure on G^n (which is also the stationary measure on G^n). The following is our main theorem. We remark that although the theorem is given in its most general form, the special case where G is a regular graph is already interesting. For this special case μ is simply the uniform measure.

Theorem 1.1. *Let $G = (V, E)$ be a simple, undirected, connected, non-bipartite graph. Then there exists a function $j = j(\varepsilon)$ such that if $I \subseteq V^n$ is an independent set in G^n then for every $\varepsilon > 0$ there exists a set $J \subseteq V^n$ depending on at most j coordinates such that $\mu(I \setminus J) \leq \varepsilon$ and such that J spans less than $\varepsilon|E(G^j)|$ edges in the graph G^j.*

Remarks:

- If G is either bipartite or not connected, then there is no unique stationary measure, and it is not hard to see that the assertion of the theorem is false for the uniform measure.
- Our proof shows that the theorem also holds for the case that I is a sparse set rather than an independent set. More precisely, there is a function $\zeta(\varepsilon) > 0$, such that for all $\varepsilon > 0$, if $I \subseteq V^n$ is a set spanning less than $\zeta(\varepsilon)|E(G^n)|$ edges in G^n then there exists a set $J \subseteq V^n$ as in Theorem 1.1.
- We wonder whether it is possible to strengthen our theorem and prove the existence of such a set J which is a bona fide independent set (rather than a sparse set) or whether this setback reflects a necessary caveat.
- The theorem and its proof can be easily extended to multigraphs (graphs G with multiple edges and self-loops) or, more generally, to graphs with weighted edges. Equivalently, we can think of G as a reversible (finite, aperiodic, irreducible) Markov chain. So in Theorem 1.1 the measure μ becomes the stationary measure of the random walk on the weighted graph, and the number of edges spanned by J is now replaced with the weight of edges spanned by J. For more details, see [9] where this extension is applied to deduce a theorem concerning the structure of intersecting families of sets.
- For which connected, non-bipartite graphs does the assertion of the theorem hold under the *uniform* measure? Here we note that these are exactly the *regularizable* graphs. A graph is called regularizable if a regular graph can be obtained from it by replacing each edge by some positive number of edges. By the previous remark, it follows that for any connected, non-bipartite, regularizable graph, the assertion of the theorem holds under the uniform measure. For the converse, we use a characterization of Berge [5], which says that a connected, non-bipartite graph is regularizable if and only if every non-empty independent set $S \subseteq V$ has more neighbors than elements, i.e., $|N(S)| > |S|$. Now note that if there exists a non-empty independent set S such that $|N(S)| \leq |S|$, then the assertion of the theorem is false under the uniform measure since we can consider the independent set in G^n given by all vertices in which more than $n|S|/|V|$ coordinates are from S and less than $n|S|/|V|$ coordinates are from $N(S)$.

The natural emergence of cylinders in graph powers appears already in [13] where it was proven that for a certain class of base graphs G the only optimal colorings are those induced by a coloring of the base graph. This was reproved in [1] via an analytical approach, which also gave the following robustness theorem: for a wide family of regular base graphs G any independent set in G^n whose measure is $1 - \varepsilon$ times the maximum value, is $O(\varepsilon)$ near to a cylinder over an independent set in the base graph. Furthermore, [1] conjectured that for regular base graphs every independent set whose size is $\alpha|V|^n$, for any constant $\alpha < 1$, is contained (up to a negligible error) in a junta. This is essentially a special case of what we prove in this paper.

Our main new tool here that was absent in [1] is a corollary of the powerful new invariance principle of Mossel, O'Donnell and Oleszkiewicz as presented in [18]. Their approach gives the basic leverage needed for our proof: the fact that when two sets of vertices have few edges between them then they can be described by functions which depend jointly, in a non-negligible manner, on a certain coordinate. This idea was previously exploited in [11], a forerunner of this paper, where it was used to derive certain inapproximability results for coloring problems. (This was one of the original motivations for studying these questions in the first place.) Some other tools that arise in the proof are a hypercontractive inequality due to Miclo [17], and the notion of fractional expansion in a graph, first described in [2].

2 Background: The Spectral Basis, Influences, Noise and Hypercontractivity

2.1 An Analog of Fourier Expansion

Let $G = (V, E)$ be a simple, undirected, connected and non-bipartite graph. We write $u \sim v$ to mean that the vertices $u, v \in V$ are adjacent in G. Let $V = \{0, 1, \dots, |V| - 1\}$.

We will consider independent sets in G^n as functions $f : V^n \to \{0, 1\}$. We will also represent these functions as column vectors and let matrices representing linear operators operate on them from the left. In the spirit of [1], we will consider the following analog of the Fourier expansion. Just as the Fourier-Walsh basis for the space of functions on $\{0, 1\}^n$ is an n-fold tensor product of a two-dimensional basis, we will define a similar basis for functions $f : V^n \to \mathbb{R}$. This basis will enjoy many of the nice properties of the Fourier-Walsh basis.

We begin by defining a basis for the space of functions from V to $\mathbb{R}$. Let A be the transition matrix of the simple random walk on G (when acting *from the right* on row vectors), henceforth *the transition matrix of* G:

$$A_{u,v} = \begin{cases} 1/\deg(u) & u \sim v \\ 0 & \text{otherwise}\,. \end{cases}$$

Let μ be the unique stationary measure of this walk on V, henceforth the *stationary measure*,

$$\mu(u) = \frac{\deg(u)}{2|E|}.$$

Let $\mathcal{M}$ denote the diagonal $|V| \times |V|$ matrix with entries $\mathcal{M}_{u,v} = \delta_{u,v}\mu(u)$. We will consider the inner product on $\mathbb{R}^V$ defined by

$$\langle f, g\rangle_\mu = \mathbb{E}_\mu[f(u)g(u)] = \sum_{u \in V} f(u)g(u)\mu(u) = f^t \mathcal{M} g.$$

Also, for any $p \geq 1$, the p-th norm of f is defined as $\|f\|_p = (\mathbb{E}_\mu|f|^p)^{1/p}$. Notice that $\|f\|_2^2 = \langle f, f\rangle_\mu$. The following lemma states some standard basic facts related to such a Markov chain, see, e.g., [16].

Lemma 2.1. *There exists a basis* $\{\chi_0, \chi_1, \dots \chi_{|V|-1}\}$ *for* $\mathbb{R}^V$*, which consists of (right) eigenvectors of* A *corresponding to a set of real eigenvalues*

$$1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|V|-1} > -1 \tag{1}$$

such that the basis is orthonormal with respect to the above inner product, i.e.,

$$\langle \chi_i, \chi_j \rangle_\mu = \delta_{i,j}.$$

Furthermore, $\chi_0(u) = 1$ *for all* u *in* V.

Proof: The statement concerning the range of A's eigenvalues follows from the fact that A represents a reversible, irreducible, aperiodic Markov chain.

Let $B = \sqrt{\mathcal{M}} A \sqrt{\mathcal{M}^{-1}}$. It is easy to see that B is symmetric and hence has a complete set of eigenvectors $\{w_i\}_{i=0,\ldots,|V|-1}$ which are orthonormal with respect to the standard Euclidean inner product, i.e.,

$$w_i^t w_j = \delta_{i,j}.$$

Since B is similar to A they have the same eigenvalues so we can assume $Bw_i = \lambda_i w_i$. Now define $\chi_i = \sqrt{\mathcal{M}^{-1}} w_i$. This is the required basis. Indeed,

$$A\chi_i = \sqrt{\mathcal{M}^{-1}} B \sqrt{\mathcal{M}} \sqrt{\mathcal{M}^{-1}} w_i = \sqrt{M^{-1}} \lambda_i w_i = \lambda_i \chi_i$$

and

$$\langle \chi_i, \chi_j \rangle_\mu = \chi_i^t \mathcal{M} \chi_j = w_i^t \sqrt{\mathcal{M}^{-1}} \mathcal{M} \sqrt{\mathcal{M}^{-1}} w_j = \delta_{i,j}.$$

In particular the fact that A is (right) stochastic implies that $\chi_0 \equiv 1$. ■

We now proceed to tensor the above basis. This yields a basis of eigenvectors of $A^{\otimes n}$ given by all

$$\chi_S = \bigotimes_{j=1}^{n} \chi_{S_j} \tag{2}$$

as S ranges over all multi-indices $S = (S_1, \ldots, S_n) \in V^n$. This basis is an orthonormal basis for the space of real valued functions on V^n with respect to the stationary measure $\mu^{\otimes n}$, which will also be denoted as μ. The corresponding eigenvalues are

$$\lambda_S = \prod_{j=1}^{n} \lambda_{S_j},$$

and we also define

$$|S| = |\{j : \; S_j \neq 0\}|.$$

A key notion in what follows is the *spectral gap* of G and G^n. Let

$$\lambda(A) = \lambda(G) = \max\{|\lambda_i| : \; i \neq 0\}.$$

In the cases that we consider $\lambda(A)$ will always be strictly less than 1. Note that for any $n \geq 1$ we have $\lambda(G^n) = \lambda(G)$.

Adopting the usual Fourier notation we write the expansion of $f : V^n \to \mathbb{R}$ according to the tensor basis as

$$f = \sum \hat{f}(S) \chi_S.$$

From the orthogonality of the basis vectors we have

$$\langle f, g \rangle_\mu = \sum \hat{f}(S) \hat{g}(S)$$

and in particular $\|f\|_2^2 = \sum \hat{f}(S)^2$.

2.2 Influence

Since we are interested in proving that certain functions essentially depend on few coordinates we need a way to measure this dependence. The greatly influential notion of the *influence of a variable* on a function f whose domain is a product space was introduced by Ben-Or and Linial in [4]. They measured the probability that when one chooses a point in the domain of f at random, the value of the function changes if the ith variable changes its value. Let us make this precise in our case. Let f be a real valued function on V^n. Define a function $f_i : V^n \to \mathbb{R}$ by

$$f_i(x) = f(x) - \mathbb{E}_{\mu,x_i}[f(x)] \, .$$

For example, in the case of the graph $(K_3)^n$, $f_i(x) = \frac{2f(x)-f(x+e_i)-f(x+2e_i)}{3}$, where e_i is the vector $(0, 0, \ldots, 1, \ldots, 0)$ with 1 in the ith coordinate, and addition is taken modulo 3.

Clearly f_i measures the dependence of f on the ith variable. We will define the influence of the ith variable on f by

$$\mathrm{Inf}_i(f) = \|f_i\|_2^2.$$

A nice aspect of this definition, as observed, e.g., in [15], is that it has a very simple expression in terms of the Fourier expansion of f. It is not hard to see that

$$\widehat{f_i}(S) = \begin{cases} \hat{f}(S) & S_i \neq 0 \\ 0 & S_i = 0 \end{cases} .$$

Therefore

$$\mathrm{Inf}_i(f) = \sum_{S:S_i \neq 0} \hat{f}(S)^2.$$

As mentioned in the introduction, a crucial tool in this paper is a result stemming from the work of Mossel, O'Donnell and Oleszkiewicz [18]. The importance of this tool is that it enables us to find a variable that has large influence simultaneously on two (and ultimately on many) different functions. This will be crucial when trying to identify variables participating in a junta. The statement below is basically derived from [18] in [11, Theorem 3.1], where it is used for similar purposes.[1]

Theorem 2.2 (Corollary of Invariance Principle)**.** *Let G be a connected, non-bipartite graph with transition matrix A and let μ be its stationary measure. Then, there exist functions $\delta = \delta_{\mathrm{MOO}}(\varepsilon) > 0$ and $\tau = \tau_{\mathrm{MOO}}(\varepsilon) > 0$ such that for any $\varepsilon > 0$, $n \geq 1$, and functions $g_1, g_2 : V^n \to [0,1]$ with $\mathbb{E}_\mu[g_1] \geq \varepsilon$, $\mathbb{E}_\mu[g_2] \geq \varepsilon$ and $\langle g_1, Ag_2 \rangle_\mu < \delta$, there exists a coordinate i with influence τ on both functions, i.e.,*

$$\mathrm{Inf}_i(g_1) > \tau \ \text{ and } \ \mathrm{Inf}_i(g_2) > \tau.$$

2.3 The Noise Operator

A central tool in the application of discrete Fourier analysis on product spaces are noise operators (see, e.g., [7, 10]) and these will play a crucial role in this paper too. Fix some set V and let μ be some probability measure on V. For any $0 \leq \eta \leq 1$, we define the noise operator N_η on the space of functions on V by

$$(N_\eta f)(x) = \eta f(x) + (1-\eta)\mathbb{E}_\mu[f].$$

[1] There are two small technical differences. First, [11] only considers the special case where μ is uniform, but their proof easily generalizes to our case. Second, the conclusion of [11, Theorem 3.1] regards the k-degree influence (see ibid.) as opposed to the influence. Since the former is a lower bound on the latter, our statement follows.

Clearly N_1 is the identity operator whereas N_0 is the averaging operator, mapping each f to the constant function $\mathbb{E}_\mu[f]$. It is often useful to think of N in the following way. For $x \in V$, $(N_\eta f)(x)$ is the expectation of $f(y)$ where y is chosen to be x with probability η and chosen according to μ with probability $1-\eta$.

The eigenvalues and eigenvectors of N_η have a particularly simple description: the constant function is an eigenvector with eigenvalue 1, whereas the orthogonal subspace containing all functions whose expectation under μ is zero has eigenvalue η. In particular, if G is a connected non-bipartite graph on vertex set V, with stationary measure μ, then its eigenvectors $\chi_0, \chi_1, \ldots, \chi_{|V|-1}$, also form a complete set of eigenvectors of N_η with corresponding eigenvalues $1, \eta, \ldots, \eta$.

We now tensor the noise operator to obtain $(N_\eta)^{\otimes n}$. For any $x \in V^n$ and $f : V^n \to \mathbb{R}$, $((N_\eta)^{\otimes n} f)(x)$ can be seen as the expectation of $f(y)$ where each coordinate y_i of y is chosen, independently of the other coordinates, to be x_i with probability η and otherwise it is chosen according to μ. For any graph G as before, $\{\chi_S\}_{S \in V^n}$ forms a complete set of eigenvectors of $(N_\eta)^{\otimes n}$ with corresponding eigenvalues $\eta^{|S|}$. This decay of the "higher frequencies" helps to explain why the noise operator has a smoothing effect on functions. This will also be true for $A^{\otimes n}$, as described in Theorem 2.6 (see also [21]).

The following lemma summarizes some properties of the noise operator. By abuse of notation, we sometimes write N_η to denote $(N_\eta)^{\otimes n}$ or $(N_\eta)^{\otimes(n-j)}$ etc., and similarly for A.

Lemma 2.3. *Let $\eta \in [0,1]$ and $f : V^n \to \mathbb{R}$. Then,*

1. **Noise preserves mass.** *If $f \geq 0$ then $\|f\|_1 = \|N_\eta f\|_1$.*
2. **Noise preserves monotonicity.** *If $f \geq f^*$ pointwise then $N_\eta f \geq N_\eta f^*$ pointwise.*
3. **Noise decreases the range.** *If $f(x) \in [a,b]$ for all $x \in V^n$ then also $N_\eta f(x) \in [a,b]$ for all $x \in V^n$.*
4. **Noise commutes with averaging.** *Let $j \geq 0$ and for $v \in V^n$ write $v = (a,x)$ with $a \in V^j$ and $x \in V^{n-j}$. Then,*
$$\mathbb{E}_{\mu,x}[(N_\eta)^{\otimes n} f(a,x)] = (N_\eta)^{\otimes j} \mathbb{E}_{\mu,x}[f(a,x)].$$
5. **Noise decreases influences.** $\mathrm{Inf}_i(N_\eta f) \leq \mathrm{Inf}_i(f)$ *for all $i \in \{1, \ldots, n\}$.*
6. **The sum of influences of a noisy version of a bounded function is bounded.** *If $|f(x)| \leq 1$ for all $x \in V^n$ then,*
$$\sum_{i=1}^n \mathrm{Inf}_i(N_\eta f) \leq (1-\eta^2)^{-2}.$$
In particular, the number of variables in $N_\eta f$ that have influence larger than τ is at most $\frac{(1-\eta^2)^{-2}}{\tau}$.

Proof: Items 1, 2 and 3 follow from the fact that N_η is an averaging operator. For Item 4 it suffices to verify the equality on the basis vectors $\{\chi_S\}_S$. It is easy to see that if $S \subseteq [j]$ then both sides are $\eta^{|S|}\chi_S$ (when viewed as a function on V^j) and if $S \nsubseteq [j]$ then both sides are zero. Item 5 follows directly from the formula $\mathrm{Inf}_i(f) = \sum_{S: S_i \neq 0} \hat{f}^2(S)$, and because for all S

$$|\widehat{N_\eta f}(S)| = |\eta^{|S|}\hat{f}(S)| \leq |\hat{f}(S)|.$$

For Item 6, notice that $|f| \leq 1$ implies that $\sum \hat{f}(S)^2 = \|f\|_2^2 \leq 1$. So,

$$\sum_{i=1}^n \mathrm{Inf}_i(N_\eta f) = \sum_{i=1}^n \sum_{S: S_i \neq 0} (\widehat{N_\eta f}(S))^2 = \sum_S |S| \eta^{2|S|} (\hat{f}(S))^2 < \sum_{k=1}^\infty k\eta^{2k} \leq (1-\eta^2)^{-2}.$$

■

2.4 The Transition Matrix of G^n

We now consider the transition matrix of G^n in more detail. We first notice, following [1], that this matrix is useful for identifying independent sets in G^n.

Observation 2.4. *Let $J \subseteq V^n$. The expression $\langle \mathbf{1}_J, A\mathbf{1}_J \rangle_\mu$ is equal to the fraction of the edges of G^n that are spanned by J. Consequently, J is an independent set in G^n if and only if $\langle \mathbf{1}_J, A\mathbf{1}_J \rangle_\mu = 0$.*

Proof:

$$\langle \mathbf{1}_J, A\mathbf{1}_J \rangle_\mu = \sum_u \mu(u)\mathbf{1}_J(u)A\mathbf{1}_J(u) = \sum_u \frac{\deg(u)}{2|E|}\mathbf{1}_J(u) \sum_{v:\, v\sim u} \mathbf{1}_J(v)/\deg(u) = \frac{1}{|E|}\sum_{u\sim v}\mathbf{1}_J(u)\mathbf{1}_J(v)\,.$$

■

Next, we observe that the expression $\langle \mathbf{1}_J, A\mathbf{1}_J \rangle_\mu$ does not change much if we replace $\mathbf{1}_J$ with an η-noisy version of it when η is close to 1.

Lemma 2.5. *Let $\lambda = \lambda(G) = \lambda(A)$. Let $1 - \lambda < \eta < 1$ be sufficiently close to 1 so that*

$$(1-\eta)\log_\lambda(1-\eta) \le \sqrt{1-\eta}. \tag{3}$$

Let $g = N_\eta f$. If $|f(x)| \le 1$ for all x, then

$$\left|\langle f, Af \rangle_\mu - \langle g, Ag \rangle_\mu\right| \le \sqrt{1-\eta}.$$

Proof: For all S we have $\hat{g}(S) = \eta^{|S|}\hat{f}(S)$, so

$$\left|\langle g, Ag \rangle_\mu - \langle f, Af \rangle_\mu\right| = \left|\sum_S \left(\hat{g}(S)^2 - \hat{f}(S)^2\right)\cdot\lambda_S\right| \le \sum_S \hat{f}(S)^2(1-\eta^{2|S|})\cdot|\lambda_S|.$$

Since $\sum_S \hat{f}(S)^2 = \|f\|_2^2 \le 1$ our result will follow if we show that for all S

$$(1-\eta^{2|S|})\cdot|\lambda_S| \le \sqrt{1-\eta}.$$

Let $r = \frac{1}{2}\log_\lambda(1-\eta)$. Then for $|S| \le r$ we have

$$1-\eta^{2|S|} \le 2|S|(1-\eta) \le 2r(1-\eta)$$

which, by our choice of r and by (3) is at most $\sqrt{1-\eta}$. For $|S| > r$ we have by the definition of r and of λ

$$|\lambda_S| \le \lambda^{|S|} \le \sqrt{1-\eta}.$$

■

2.5 Hypercontractivity and Expansion

A key feature of the graph G^n that we use is its strong expansion properties. As we will show below, these properties follow from the fact that $A^{\otimes n}$, the normalized adjacency matrix of G^n, satisfies what is known as a *hypercontractive inequality*. Such an inequality bounds some high norm of $A^{\otimes n}f$ by some low norm of f, for any $f : V^n \to \mathbb{R}$ and $n \ge 1$ (see Theorem 2.6).

The intuitive connection between hypercontractive inequalities and expansion is clear: $A^{\otimes n}$ replaces $f(x)$ with the average of f over the neighbors of x. If the graph has a strong expansion, one can expect this to "smooth" the function f, thereby decreasing the larger norms. We will actually prove the reverse implication and deduce expansion properties from hypercontractivity.

2.5.1 Hypercontractivity

Hypercontractive inequalities for functions on discrete product spaces have been proven independently by several different people: Bonami [6], Gross [14], Beckner [3], and possibly others. These inequalities have become an indispensable tool in the study of Boolean functions, see, e.g., [15, 12, 1, 18, 19]. It is also of interest to find the optimal constants in such inequalities, see, e.g., [20] and [22].

It is a well known fact that there is a connection between the log-Sobolev constant of a Markov chain and the hypercontractivity of the associated continuous semi-group see, e.g., [8]. However, we need the hypercontractivity of the transition matrix A of the discrete (finite) Markov chain. Precisely such a result is given in [17] by Miclo, who uses the the log-Sobolev constant of A itself to derive its hypercontractivity.[2]

Theorem 2.6 (Hypercontractive Inequality, [17])**.** *Let G be a connected, non-bipartite graph with transition matrix A. Then there is a constant $p = p(G) > 2$ which does not depend on n, such that for all functions $f : V^n \to \mathbb{R}$*

$$\|Af\|_p \leq \|f\|_2,$$

where the norms are with respect to the stationary measure of the random walk on G^n.

As usual in proving hypercontractive inequalities concerning operators on product spaces, it suffices to treat the one dimensional case due to the submultiplicativity of the operator norms (see, e.g., [6, 3]). Hence the fact that A is hypercontractive implies that $A^{\otimes n}$ too is hypercontractive with the same constants.

In a recent preprint [22] Wolff finds the optimal constants for hypercontractivity of such operators. For the sake of self containedness we present in the appendix a loose adaptation of certain portions of his proof with no attempt to compute any constants, making do with proving the existential statement of Theorem 2.6.

2.5.2 Expansion and Fractional Expansion

The usual notion of expansion for a graph G implies that for any set X of vertices in G, the size of $\Gamma(X)$, its neighbor set, is at least of the same order of magnitude as the size of X. In other words, if for each $x \in X$ we let Γ_x be the set of neighbors of x, then for $\Gamma = \Gamma(X) := \cup_{x \in X} \Gamma_x$ we have $|\Gamma| = \Omega(|X|)$. Using the hypercontractivity of A, we will see in Lemma 2.7 below that G^n satisfies an expansion property which, for small sets X, is much stronger, namely, that $\mu(\Gamma) \geq \mu(X)^{2/p}$ where $p > 2$ is the constant from Theorem 2.6.

In fact, for our application we need to prove an even stronger property known as *fractional expansion*, a notion which first appeared in [2]. Let $0 < \beta \leq 1$ be a constant, and let B_x be an arbitrary subset of Γ_x, containing at least a $\beta > 0$ fraction of x's neighbors. Fractional expansion says that for any such setting, the set $B = \cup_{x \in X} B_x$ is still much larger than the set X. In our lemma below we actually prove a slightly stronger statement by only requiring that the *average* fractional size of B_x is at least β. This lemma and its proof are adapted from the similar statement and proof of Theorem 6 in [2].

Lemma 2.7. *Let $G = (V, E)$ be a connected non-bipartite graph and let $p = p(G)$ the constant defined in Theorem 2.6. Let $j \geq 1$ and $X \subseteq V^j$ and let $0 < \beta < 1$. For each $x \in X$ let B_x be a set of neighbors of x such that*

$$\sum_{x \in X} |B_x| \geq \beta \sum_{x \in X} \deg_{G^j}(x).$$

[2] In fact, Miclo's result is far more general, and applies to arbitrary Markov chains under some mild conditions.

Let $B = \cup_{x \in X} B_x$. Then

$$\mu(B) \geq \mu(X)^{\frac{2}{p}} \beta^2$$

where, as usual, μ denotes the stationary measure of the random walk on G^j.

Proof: Let $\mathbf{1}_B$ be the indicator function of B. Then our assumption implies that

$$\sum_{x \in X} \mu(x) \cdot (A\mathbf{1}_B)(x) \geq \mu(X)\beta.$$

By convexity it follows that

$$\|A\mathbf{1}_B\|_p^p \geq \sum_{x \in X} \mu(x) \cdot ((A\mathbf{1}_B)(x))^p \geq \mu(X)\beta^p.$$

Hence, by the hypercontractive inequality of Theorem 2.6,

$$\|\mathbf{1}_B\|_2^p \geq \mu(X)\beta^p.$$

But $\|\mathbf{1}_B\|_2^p = \mu(B)^{\frac{p}{2}}$ so we are done. ■

The following lemma is a consequence of the strong fractional expansion property of G^j. It shows that if we assign to each vertex an arbitrary list of at most ℓ labels in such a way that for at least ε of the edges of G^j the lists corresponding to their two endpoints have a nonempty intersection, then there must exist a "popular" label contained in many of the lists. It can be interpreted as saying that some weak local consistency implies certain global consistency.

Lemma 2.8. *Let G and $p = p(G)$ be as above. Let $\varepsilon > 0$, $\ell, j \in \mathbb{N}$ be parameters. For every vertex $a \in V^j$ let $L(a) \subseteq \mathbb{N}$ be a set of labels with $|L(a)| \leq \ell$. Assume also that for at least an ε fraction of the edges $\{a, b\}$ in G^j it holds that $L(a) \cap L(b) \neq \emptyset$. Then there exists a label $i \in \mathbb{N}$ such that*

$$\mu(\{a \,:\, i \in L(a)\}) \geq (\varepsilon/\ell^2)^{\frac{2p}{p-2}}.$$

Proof: Consider the labelling of V^j obtained by choosing for each vertex one random label from its label set. Then, for each edge $\{a, b\}$ satisfying $L(a) \cap L(b) \neq \emptyset$, the probability that both its endpoints obtain the same label is at least $1/\ell^2$. Hence, it follows that there exists a labelling such that the set of edges H both of whose endpoints have the same label is of size at least $\frac{\varepsilon}{\ell^2}|E(G^j)|$. Equivalently, we have that

$$\sum_{x \in V^j} \deg_H(x) \geq \frac{\varepsilon}{\ell^2} \sum_{x \in V^j} \deg_{G^j}(x).$$

By partitioning these sums according to the connected components of H, we obtain that there exists a connected component X of H such that

$$\sum_{x \in X} \deg_H(x) \geq \frac{\varepsilon}{\ell^2} \sum_{x \in X} \deg_{G^j}(x).$$

We can now apply Lemma 2.7 with the set X, each B_x chosen to be the set of neighbors of x in H, and $\beta = \frac{\varepsilon}{\ell^2}$. By our choice of B_x, we see that $B = X$. Therefore,

$$\mu(X) \geq \mu(X)^{\frac{2}{p}} \beta^2$$

from which it follows that

$$\mu(X) \geq \beta^{\frac{2p}{p-2}} = (\varepsilon/\ell^2)^{\frac{2p}{p-2}}.$$

It remains to notice that since X is a connected component, all vertices in X must have the same label. ■

3 Proof of the Main Theorem

In this section we prove our main theorem:

Theorem 1.1. *Let $G = (V, E)$ be a simple, undirected, connected, non-bipartite graph. Then there exists a function $j = j(\varepsilon)$ such that if $I \subseteq V^n$ is an independent set in G^n then for every $\varepsilon > 0$ there exists a set $J \subseteq V^n$ depending on at most j coordinates such that $\mu(I \setminus J) \leq \varepsilon$ and such that J spans less than $\varepsilon|E(G^j)|$ edges in the graph G^j.*

In order to better understand the proof strategy let us return to a problematic example of a large independent set. Let $I_0 \subseteq V$ be some independent set in G, and let $J = I_0 \times V^{n-1}$ be the set of all vectors whose first coordinate is in I_0. Now let K be a random subset of J where every point is taken independently with probability $\frac{1}{2}$. If f is the characteristic function of K then with high probability all variables have influence of order $\Theta(1)$ on f. However, suppose we introduce a slight noise and consider $g = N_\eta f$ for some η very close to 1. Recall that, by definition, $g(x)$ is the average of f over vectors y whose first coordinate is the same as x with probability at least η. Therefore, for most points $x \in J$ the value of $g(x)$ will be very close to $\frac{1}{2}$, whereas for most points $x' \notin J$, $g(x')$ will be very close to 0. We have "recovered" J, a junta containing K. The only variable that has non-negligible influence on g is the first one, and hence when we partition V^n according to it we discover J.

The plan of the proof is now clear. Given an independent set $I \subseteq V^n$ let f be its characteristic function and take $g = N_\eta f$ to be a slightly noisy version of f. There will be few variables that have non-negligible influence on g, without loss of generality assume these are the variables indexed $1, \ldots, j$. Partition V^n into $|V|^j$ subcubes according to these variables, let $J' \subseteq V^j$ be the set of subcubes where g has non-negligible expectation, and let $J = J' \times V^{n-j}$. We will show that

- J almost contains I, that is, the expectation of f outside of J is small.
- J is almost independent, i.e., $\langle \mathbf{1}_J, A\mathbf{1}_J \rangle_\mu$ is small. In particular J is not trivially equal to V^n.

3.1 The Proof

Let us begin by setting parameters. Let $p = p(G)$ be the constant defined in Theorem 2.6 and fix some $\varepsilon > 0$. Let τ and δ be defined by

$$\tau = \tau_{\text{MOO}}(\varepsilon),\ \delta = \delta_{\text{MOO}}(\varepsilon).$$

Let $\lambda = \lambda(G)$. We choose $\eta < 1$ close enough to 1 so that $\eta > 1 - \lambda$, $\sqrt{1-\eta} \leq \frac{\delta\varepsilon}{2}$, and (3) holds. Moreover, we define

$$\ell = \frac{(1-\eta^2)^{-2}}{\tau}.$$

Finally, choose $\gamma > 0$ small enough so that

$$\gamma < \tau \cdot (\varepsilon/2\ell^2)^{\frac{2p}{p-2}}.$$

Let $I \subseteq V^n$ be an independent set, let $f : V^n \to \{0, 1\}$ be its characteristic function, and define

$$g = N_\eta f.$$

By Item 3 of Lemma 2.3 we have $g : V^n \to [0, 1]$. Let j be the number of variables with influence larger than γ on g, and assume without loss of generality that these are the variables $\{1, \ldots, j\}$. By

Item 6 of Lemma 2.3 j is upper bounded by a constant independent of n, namely $j \le \frac{(1-\eta^2)^{-2}}{\gamma}$. We will now think of V^n as $V^j \times V^{n-j}$. We will denote the elements of V^j by the letters a and b, and the elements of V^n or V^{n-j} by x and y. For every $a \in V^j$ define a function $g_a : V^{n-j} \to [0,1]$ by

$$g_a(x) = g(a, x).$$

Let

$$J' = \{a : \mathbb{E}_\mu[g_a] \ge \varepsilon\} \subseteq V^j.$$

Define the set $J = J' \times V^{n-j} \subseteq V^n$. We claim that J is the set guaranteed by Theorem 1.1. To prove this we must prove the following two claims.

Claim 3.1. $\mu(I \setminus J) \le \varepsilon$.

Claim 3.2. $\langle \mathbf{1}_{J'}, A\mathbf{1}_{J'} \rangle_\mu \le \varepsilon$.

Proof of Claim 3.1: Define $\tilde{f}(a) = \mathbb{E}_{\mu,x}[f(a,x)]$ and similarly $\tilde{g}(a) = \mathbb{E}_{\mu,x}[g(a,x)]$ the functions on V^j obtained by averaging over the V^{n-j} part. Since N_η commutes with averaging (see Item 4 of Lemma 2.3), we have $\tilde{g} = N_\eta \tilde{f}$. With this notation we have $J' = \{a : \tilde{g}(a) \ge \varepsilon\}$ and we define $f^* = \tilde{f} \cdot \mathbf{1}_{\overline{J'}}$. Notice that $\mu(I \setminus J) = \mathbb{E}_\mu[f^*]$ and hence our goal is to prove that $\mathbb{E}_\mu[f^*] \le \varepsilon$. By Item 2 of Lemma 2.3 the fact that $f^* \le \tilde{f}$ pointwise implies that $N_\eta f^* \le \tilde{g}$ pointwise. Hence we see that

$$\langle f^*, N_\eta f^* \rangle_\mu \le \langle f^*, \tilde{g} \rangle_\mu \le \varepsilon \cdot \mathbb{E}_\mu[f^*].$$

On the other hand,

$$\langle f^*, N_\eta f^* \rangle_\mu = \sum_S \widehat{f^*}(S)^2 \eta^{|S|} \ge \widehat{f^*}(\mathbf{0})^2 = (\mathbb{E}_\mu[f^*])^2.$$

Combining these two yields the required result. ∎

Proof of Claim 3.2: Let us assume by contradiction that $\langle \mathbf{1}_{J'}, A\mathbf{1}_{J'} \rangle_\mu > \varepsilon$, i.e., that the graph spanned by J' in G^j spans more than $\varepsilon|E(G^j)|$ edges. We will show that this implies that there exists a variable $i \in \{j+1, \ldots, n\}$ whose influence on g is greater than γ, in contradiction to our definition of j. To this end, notice that the influences of g are related to those of g_a by

$$\forall i \in \{j+1, \ldots, n\}, \qquad \mathrm{Inf}_i(g) = \sum_{a \in V^j} \mu(a) \mathrm{Inf}_i(g_a).$$

Therefore, in order to reach a contradiction it suffices to find a variable i with $\mathrm{Inf}_i(g_a) > \tau$ for a set of a's of measure at least $\frac{\gamma}{\tau}$.

For each $a \in V^j$ let $L(a)$ be the set of variables whose influence on g_a is greater than τ,

$$L(a) = \{j < i \le n \,:\, \mathrm{Inf}_i(g_a) > \tau\}.$$

Below we will show that

1. $|L(a)| \le \ell$ for all $a \in V^j$ (see Claim 3.3 below), and that
2. for at least $\frac{\varepsilon}{2}|E(G^j)|$ edges $\{a, b\}$ in G^j we have $L(a) \cap L(b) \ne \emptyset$ (see Claim 3.4 below).

From Lemma 2.8 it will follow that there exists an $i \in \{j+1, \ldots, n\}$ for which $\mu(\{a \,:\, i \in L(a)\}) \ge (\varepsilon/2\ell^2)^{\frac{2p}{p-2}}$. Recalling that we have chosen parameters so that $(\varepsilon/2\ell^2)^{\frac{2p}{p-2}} > \frac{\gamma}{\tau}$ we reach the contradiction $\mathrm{Inf}_i(g) > \gamma$ and our argument is complete. It remains to prove the claims in the two items above.

Claim 3.3. *For all $a \in V^j$, $\sum \mathrm{Inf}_i(g_a) \leq (1-\eta^2)^{-2}$. In particular, there are at most $\ell = (1-\eta^2)^{-2}/\tau$ variables whose influence on g_a is at least τ.*

Proof: Let $h : V^n \to [0,1]$ be the result of applying noise on the first j coordinates of f,

$$h = (N_\eta)^{\otimes j} \otimes I^{\otimes n-j} f .$$

For each $a \in V^j$ let $h_a : V^{n-j} \to [0,1]$ be defined by $h_a(x) = h(a,x)$. The function g_a can now be written as $g_a = N_\eta h_a$. By Item 6 of Lemma 2.3, $\sum \mathrm{Inf}_i(g_a) \leq (1-\eta^2)^{-2}$. ■

For the second claim, we say that an edge $\{a,b\}$ in G^j is *dense* if $\langle g_a, Ag_b \rangle_\mu \geq \delta$.

Claim 3.4. *There are at most $\frac{\varepsilon}{2}|E(G^j)|$ dense edges in G^j.*

Proof: First note that since $\langle f, Af \rangle_\mu = 0$, and since $g = N_\eta f$, Lemma 2.5 says that

$$\langle g, Ag \rangle_\mu \leq \sqrt{1-\eta}.$$

Next, write

$$g = \sum_a \mathbf{1}_{\{a\}} \otimes g_a,$$

and deduce that

$$\langle g, Ag \rangle_\mu = \frac{1}{|E(G^j)|} \sum_{a \sim b} \langle g_a, Ag_b \rangle_\mu.$$

If there were more than $\frac{\varepsilon}{2}|E(G^j)|$ dense edges they would contribute together more than $\delta\frac{\varepsilon}{2} \geq \sqrt{1-\eta}$ to $\langle g, Ag \rangle_\mu$. ■

Consider now all non-dense edges spanned by J'. By our hypothesis and the claim above, there are at least $\frac{\varepsilon}{2}|E(G^j)|$ such edges. Moreover, by definition each such edge $\{a,b\}$ satisfies $\langle g_a, Ag_b \rangle_\mu < \delta$, $\mathbb{E}_\mu[g_a] \geq \varepsilon$, and $\mathbb{E}_\mu[g_b] \geq \varepsilon$. By Theorem 2.2 this means that there exists a variable in $\{j+1, \ldots, n\}$ whose influence on both g_a and g_b is larger than τ, which implies that $L(a) \cap L(b) \neq \emptyset$, thereby completing the proof. ■

Acknowledgments

We wish to thank the anonymous referee for various useful suggestions. We would also like to thank Frank Barthe, Laurent Miclo, Krzysztof Oleszkiewicz, Prasad Tetali and Pawel Wolff for sharing with us information regarding log-Sobolev inequalities and hypercontractivity.

References

[1] N. Alon, I. Dinur, E. Friedgut, and B. Sudakov. Graph products, Fourier analysis and spectral techniques. *GAFA*, 14(5):913–940, 2004.

[2] N. Alon, G. Kalai, M. Ricklin, and L. Stockmeyer. Lower bounds on the competitive ratio for mobile user tracking and distributed job-scheduling. *Theoretical Computer Science*, 130:175–201, 1994.

[3] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, 102:159–182, 1975.

[4] M. Ben Or and N. Linial. Collective coin flipping. In S. Micali, editor, *Randomness and Computation*, pages 91–115. Academic Press, 1989.

[5] C. Berge. Regularisable graphs. I, II. *Discrete Math.*, 23(2):85–89, 91–95, 1978.

[6] A. Bonami. Etude des coefficients Fourier des fonctiones de $l^p(g)$. *Ann. Inst. Fourier (Grenoble)*, 20(2):335–402, 1970.

[7] J. Bourgain. On the distribution of the Fourier spectrum of Boolean functions. *Israel J. of Math.*, 131:269–276, 2002.

[8] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.

[9] I. Dinur and E. Friedgut. Intersecting families are essentially contained in juntas. Preprint.

[10] I. Dinur, E. Friedgut, G. Kindler, and R. O'Donnell. On the Fourier tails of bounded functions over the discrete cube. *Israel J. of Math.*, 2006. To appear. Conference version in STOC 2006.

[11] I. Dinur, E. Mossel, and O. Regev. Conditional hardness for approximate coloring. In *Proc. 38th ACM Symp. on Theory of Computing*, 2006.

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

[13] D. Greenwell and L. Lovász. Applications of product colouring. *Acta Math. Acad. Sci. Hungar.*, 25:335–340, 1974.

[14] L. Gross. Logarithmic Sobolev inequalities. *American Journal of Math.*, 97:1061–1083, 1975.

[15] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proc. 29th IEEE Symp. on Foundations of Computer Science*, pages 68–80, 1988.

[16] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Van Nostrand, 1960.

[17] L. Miclo. Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies. In *Séminaire de Probabilités, XXXI*, volume 1655 of *Lecture Notes in Math.*, pages 136–167. Springer, Berlin, 1997.

[18] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *Proc. 46th IEEE Symp. on Foundations of Computer Science*, 2005.

[19] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov. Non-interactive correlation distillation, inhomogeneous Markov chains, and the reverse Bonami-Beckner inequality. *Israel Journal of Mathematics*, 154:299–336, 2006.

[20] K. Oleszkiewicz. On a nonsymmetric version of the Khinchine-Kahane inequality. *Progress in Probability*, 56:156–168, 2003.

[21] Sade. Smooth operator. http://www.cs.tau.ac.il/∼odedr/smooth.swf.

[22] P. Wolff. Hypercontractivity of simple random variables. Preprint.

A Proof of the Hypercontractive Inequality

In this appendix we provide a proof of the hypercontractive inequality of Theorem 2.6. As mentioned before, our proof is a loose adaptation of parts of the proof by P. Wolff appearing in [22]. Although Wolff actually finds the optimal constants in this inequality we make no attempt to reproduce such a quantitative statement.

Proof: Let $m \geq 2$ be an integer and consider Ω, an m point space endowed with a probability measure μ, and the norms induced by μ on $L_p(\Omega)$ for all $p \geq 1$. Let $V_1, \ldots, V_m$ be an orthonormal basis for $L_2(\Omega)$ with the inner product induced by μ, and assume $V_1 = (1, 1, \ldots, 1)$. Let

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_m > -1$$

and define a linear operator A as follows. For $1 \leq i \leq m$ define $AV_i = \lambda_i V_i$ and extend A linearly to all $f : \{1, \ldots, m\} \to \mathbb{R}$. We want to prove the existence of $p > 2$ such that for all $(c_1, c_2, \ldots, c_m) \in \mathbb{R}^m$

$$\left\| \sum c_i \lambda_i V_i \right\|_p \leq \left\| \sum c_i V_i \right\|_2 . \tag{4}$$

First note that we may assume that $1 > \lambda_2 = \lambda_3 = \cdots = \lambda_m = \lambda > 0$. Indeed, for g in the range of A write $g = \sum d_i V_i$ (where $d_i = 0$ whenever $\lambda_i = 0$). Then (4) translates to

$$\left\| \sum d_i V_i \right\|_p \leq \left\| \sum d_i \lambda_i^{-1} V_i \right\|_2 = \left(\sum (d_i \lambda_i^{-1})^2 \right)^{1/2}$$

and the right hand side is monotone decreasing in $|\lambda_i|$. So in the following we assume that A acts by $Af = \lambda f + (1 - \lambda)\mathbb{E}_\mu[f]$.

Our goal is to show that for some $p > 2$, $\|Af\|_p \leq \|f\|_2$ holds for all $f : \{1, \ldots, m\} \to \mathbb{R}$. We first observe that it suffices to consider non-negative f. Indeed, replacing f by $|f|$ does not change $\|f\|_2$ and can only increase $\|Af\|_p$ since $A|f| \geq |Af|$ (which follows from $A|f| \geq Af$ and $A|f| \geq -Af$). Next, by homogeneity, it is enough to consider f such that $\|f\|_2 = 1$. Let

$$a := \mathbb{E}_\mu[f] = \sum \mu(i) f(i)$$

and notice that $0 < a \leq 1$.

So our goal now is to show that $\|Af\|_p \leq 1$ holds for any function f of the form

$$f = (a + b_1, a + b_2, \ldots, a + b_m)$$

for some $0 < a \leq 1$, with

$$\sum \mu(i) b_i = 0, \tag{5}$$

$$\sum \mu(i)(a + b_i)^2 = 1, \tag{6}$$

and

$$\forall i,\ a + b_i \geq 0. \tag{7}$$

To this end, define the function

$$\phi_{a,p}(b_1, \ldots, b_m) := \|Af\|_p^p = \sum \mu(i)(a + \lambda b_i)^p$$

where the equality follows from

$$Af = \lambda f + (1-\lambda)\mathbb{E}_\mu[f] = (a+\lambda b_1, a+\lambda b_2, \ldots, a+\lambda b_m).$$

In the following, we maximize $\phi_{a,p}(b_1,\ldots,b_m)$ subject to the constraints (5), (6), and (7) for any fixed a and p, and show that for sufficiently small $p>2$ and all $0<a\leq 1$ this maximum is 1.

We use the method of Lagrange multipliers. The gradient of $\phi_{a,p}$ is given by

$$\nabla\phi_{a,p}(b_1,\ldots,b_m) = \big(\lambda p\, \mu(i)(a+\lambda b_i)^{p-1}\big)_{i=1}^m . \tag{8}$$

Since this is a strictly positive vector, we see that the maximum of $\phi_{a,p}$ is not constrained by (7). Therefore, at any maximum of $\phi_{a,p}$, the gradient in (8) should be in the span of the gradients of the constraints (5) and (6). It follows that at any maximum point there exist constants t and s such that for all i

$$(a+\lambda b_i)^{p-1} = t + s b_i.$$

However, for any s,t and $p>2$ this last equation in b_i has at most two solutions. Therefore, by (5), there is some $\Omega' \subseteq \Omega$ and $b \geq 0$ such that $b_i = (1-\alpha)b$ for $i\in\Omega'$ and $b_j = -\alpha b$ for $j\notin\Omega'$ where $\alpha = \mu(\Omega')$. If $\alpha=0$ or $\alpha=1$ then f is a constant function in which case $\|Af\|_p = \|f\|_2$ for all p and we are done. Otherwise, using (6), we get $b = b(a,\alpha) = \sqrt{\frac{1-a^2}{\alpha(1-\alpha)}}$. For the function to satisfy (7), we must have $a - \alpha b \geq 0$, which simplifies to $a\geq\sqrt{\alpha}$. Define

$$\Phi_\alpha(a,p) := \phi_{a,p}(b_1,\ldots,b_m) = \alpha\left[a+\lambda\sqrt{\frac{(1-a^2)(1-\alpha)}{\alpha}}\right]^p + (1-\alpha)\left[a-\lambda\sqrt{\frac{(1-a^2)\alpha}{1-\alpha}}\right]^p .$$

So our goal in the following is to show that there exists some $p>2$ such that $\Phi_\alpha(a,p)\leq 1$ holds for all $0<\alpha<1$ and all $\sqrt{\alpha}\leq a\leq 1$. In fact, notice that there are at most 2^m-2 possible values for α, so it suffices to prove the inequality only for those values.

Fix some value of $\alpha\in(0,1)$. Notice that $\Phi_\alpha(a,2) = a^2+\lambda^2(1-a^2)$ and hence

$$\frac{\partial\Phi_\alpha(a,2)}{\partial a} = 2a(1-\lambda^2) \tag{9}$$

which is strictly positive in the range $a\geq\sqrt{\alpha}>0$. Note also that for all p

$$\Phi_\alpha(1,p) = 1. \tag{10}$$

Fix a closed rectangle, say,

$$\Delta = \big\{(a,p)\ :\ a\in[\sqrt{\alpha},1], p\in[2,3]\big\}$$

and note that both Φ_α and its derivatives are uniformly continuous on Δ. Hence we may choose $p^*_\alpha > 2$ such that for all $(a,p)\in[\sqrt{\alpha},1]\times[2,p^*_\alpha]$ we have that (9) implies

$$\frac{\partial\Phi_\alpha(a,p)}{\partial a} > 0$$

which together with (10) implies that $\Phi_\alpha(a,p)\leq 1$. Hence the assertion of the theorem holds for $p = \min_\alpha\{p^*_\alpha\} > 2$ where the minimum is taken over all finitely many possible values of α. ■