

The sparse regularity method with Schatten norms and entropy

Alexandru Pascadi^a

Submitted: Sep 12, 2024; Accepted: Jun 30, 2025; Published: Feb 13, 2026

© The author. Released under the CC BY-ND license (International 4.0).

Abstract

We introduce a regularity method for sparse graphs, with new regularity and counting lemmas which use the Schatten–von-Neumann norms to measure uniformity. This leads to k -cycle removal lemmas in subgraphs of mildly-pseudorandom graphs, and also in graphs lacking a quasi-smooth family of bipartite subgraphs, extending results of Conlon, Fox, Sudakov and Zhao. We give some additive-combinatorial applications: one about translation-invariant linear equations in subsets of mildly-pseudorandom sets, one about such equations in generalized Sidon sets, and one about polygonal patterns in subsets of \mathbb{Z}^2 with few parallelograms (giving a two-dimensional analogue for a result of Prendiville). Separately, our regularity lemma implies a dense graph removal lemma with mild constant dependencies, in graphs whose spectral $L^{2-\varepsilon}$ norms are small.

Mathematics Subject Classifications: 05C75, 11B30, 05C35

1 Introduction

A famous result in graph theory, initially motivated by the Rusza–Szemerédi problem [40] and then by its application to Roth’s theorem for 3-term arithmetic progressions [39], is the *triangle removal lemma* reproduced below.

Theorem 1 (Triangle removal lemma). *For any $\varepsilon > 0$, there exists $\delta > 0$ such that any n -vertex graph G with fewer than δn^3 triangles can be made triangle-free by removing at most εn^2 edges.*

This fact can be generalized to an H -removal lemma for any subgraph H of G [21, 22, 13], and also to the setting of hypergraphs [28, 50], the latter leading to a proof of Szemerédi’s theorem for k -term arithmetic progressions [47, 51]. We note that the known dependency of δ^{-1} on ε^{-1} (or on $\log \varepsilon^{-1}$ [22]) is at best tower-exponential, unless H belongs to a special family. The *regularity method* is the classical template used to prove such results, typically consisting of:

^aMathematisches Institut, Endenicher Allee 60, 53115 Bonn, Germany (alexpascadi@gmail.com).

1. A *regularity lemma*, which approximates G by a simpler (weighted) graph \tilde{G} ;
2. A *counting lemma*, which shows that G and \tilde{G} have about the same number of homomorphic copies of H ;
3. A *removal argument*, which removes a few blocks of edges so that \tilde{G} has no more homomorphic copies of H . After removing a few more edges, the same holds for G .

A shortcoming of results like Theorem 1 is that they are only nontrivial for *dense* graphs, which have $> \varepsilon n^2$ edges; indeed, a graph with $\leq \varepsilon n^2$ edges can always be made triangle-free by removing all edges. By contrast, a triangle removal lemma for *sparse* graphs G would assume a total of $\delta p^3 n^3$ triangles, and would remove at most $\varepsilon p n^2$ edges, where $p \in (0, 1)$ can decay to 0 as $n \rightarrow \infty$ (these values can be justified by considering the Erdős–Rényi random graph $G(n, p)$, where each edge is included independently with probability p). But such results require additional assumptions on the graphs in question, and finding appropriate assumptions has been the subject of extensive literature [17, 48, 51, 34, 33, 42, 14, 15, 1, 10, 9]. Despite their variety, most *sparse regularity methods* contain:

1. A *structure theorem / dense model theorem*, often based on a (weak) *sparse regularity lemma*. This is used to construct a *dense* graph \tilde{G} approximating the *sparse* G ;
2. A *sparse counting lemma*, which plays the same role as the dense counting lemma (but requires additional assumptions about the sparse graphs involved);
3. A *transference argument*, which applies the dense H -removal lemma to \tilde{G} , and removes a few more edges to conclude an H -removal lemma for G .

A broad category of sparse removal lemmas assume that G is a relatively-dense subgraph of a random or pseudorandom (hyper)graph, for an appropriate notion of (pseudo)randomness. Green and Tao used similar ideas to prove their celebrated theorem that the primes contain arbitrarily long arithmetic progressions [30], and Conlon, Fox and Zhao gave a simplified exposition of the proof using a more general pseudorandomness condition [17]; multidimensional analogues of this result require more restrictive (and complex) notions of pseudorandomness [48, 52]. Separately, much of the recent progress on sparse removal lemmas has focused on obtaining counting lemmas for subgraphs of random graphs, related to the resolution of the KLR conjecture [18, 5, 41].

A more recent direction of research concerns removal lemmas in sparse graphs which contain few copies of certain bipartite graphs, particularly in graphs with few 4-cycles. This was partly motivated by the work of Chung, Graham and Wilson on quasi-random graphs [11], and pursued by Conlon, Fox, Sudakov and Zhao in a series of two papers [14, 15]; their work uses a regularity lemma similar to that of Scott [42], and makes its main contribution through the counting lemmas.

The goal of this paper is to introduce a new variant of the sparse regularity method, leading to improved H -removal lemmas in both types of graphs described above, when

H is a cycle or a path; these cases also have additive-combinatorial applications about sparse subsets of abelian groups.

To highlight the difficulties faced by sparse regularity methods and to introduce our approach, we first state a version of Frieze and Kannan's weak regularity lemma [24], following the analytic form due to Lovász [36]. Here, graphs are replaced with functions $f : X \times Y \rightarrow \mathbb{C}$ (where (X, \mathbf{P}_X) and (Y, \mathbf{P}_Y) are finite probability spaces with full σ -algebras), and we will use the *cut norm* [35]

$$\|f\|_{\square} := \sup_{\substack{A \subset X \\ B \subset Y}} |\mathbf{E}_{x,y} [f(x,y) \mathbb{1}_A(x) \mathbb{1}_B(y)]|, \quad (1)$$

which has functioned in more recent literature as a substitute for the *Gowers box norm* [27, 28]

$$\|f\|_{\square^2} := \mathbf{E}_{x_1, x_2, y_1, y_2} \left[f(x_1, y_1) \overline{f(x_2, y_1)} f(x_2, y_2) \overline{f(x_1, y_2)} \right]^{1/4}. \quad (2)$$

(In the expected values from (1) and (2), it is understood that x_1, x_2, y_1, y_2 are sampled independently from the appropriate probability spaces.)

Theorem 2 (Weak regularity lemma for dense graphs [24, 36]). *Let $(X, \mathbf{P}_X), (Y, \mathbf{P}_Y)$ be finite probability spaces (with full σ -algebras), $f : X \times Y \rightarrow \mathbb{C}$, and $\varepsilon \in (0, 1)$. Then there exist partitions \mathcal{P} of X and \mathcal{Q} of Y , each with at most $2^{O(1/\varepsilon^{O(1)})}$ parts, such that*

$$\|f - \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q})\|_{\square} \leq \varepsilon \|f\|_{L^\infty}.$$

(Here we denoted $\mathcal{P} \otimes \mathcal{Q} := \{A \times B : A \in \mathcal{P}, B \in \mathcal{Q}\}$, and the conditional expectation $\mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q})$ is obtained by averaging f over parts of $\mathcal{P} \otimes \mathcal{Q}$; see Section 2.2.)

Remark 3. The reader may imagine that $f = \mathbb{1}_G/p : V \times V \rightarrow [0, \infty)$, where $\mathbb{1}_G$ is the indicator function of the edges of a graph G with vertex set V , V is equipped with the uniform probability measure, and $p = \mathbf{E}[\mathbb{1}_G]$ is roughly the edge density of G .

There are three key quantities in this theorem and its generalizations. The first is a *density norm* depending on f (in this case, the L^∞ norm), which dictates the theorem's regime of applicability, since this norm should be bounded for the conclusion to be helpful. Indeed, when $f = \mathbb{1}_G/p$ and $p = \mathbf{E}[\mathbb{1}_G]$, one has $\|f\|_{L^\infty} = 1/p$, so Theorem 2 is only useful for *dense* graphs G (with $1/p = O(1)$). The second is a *uniformity norm* (in this case, the cut norm), the smallness of which is relevant in the counting lemma. The third is the *complexity* of the partition $\mathcal{P} \otimes \mathcal{Q}$ (in this case, its cardinality); bounding this quantity makes a transference argument possible, since low-complexity functions $\mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q})$ are easier to approximate by L^∞ -bounded functions. In this context, a *sparse* regularity method may start by obtaining analogues of Theorem 2 which replace $\|\cdot\|_{L^\infty}$ with a weaker density norm, and $\|\cdot\|_{\square}$ with a stronger uniformity norm.

A key idea in our work is to use a family of norms $\|f\|_{S^q}$, for $q \in [1, \infty]$, as *both* density norms and uniformity norms. Here, $\|f\|_{S^q}$ is the Schatten q -norm (i.e., the L^q norm on the singular values) of a matrix associated to f , appropriately normalized with respect to the probability spaces; we will also refer to $\|f\|_{S^q}$ as the Schatten q -norm of f .

Using singular values to study graph regularity has of course been the topic of multiple papers [25, 23, 46, 8, 37], most of these focusing on the maximal singular value $\|f\|_{S^\infty}$. We extend Theorem 2 to the setting of these norms, with no cutoffs on f ; this requires a different notion of complexity for our partitions: Shannon entropy (see Section 2.2).

Theorem 4 (Sparse regularity lemma). *Let $(X, \mathbf{P}_X), (Y, \mathbf{P}_Y)$ be finite probability spaces (with full σ -algebras), $f : X \times Y \rightarrow \mathbb{C}$, $\varepsilon \in (0, 1)$, and $1 \leq q < r \leq \infty$. Then there exist partitions \mathcal{P} of X and \mathcal{Q} of Y , each with entropy at most $O_{q,r}(\varepsilon^{-O_{q,r}(1)})$, such that*

$$\|f - \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q})\|_{S^r} \leq \varepsilon \|f\|_{S^q}.$$

Remark 5. The crucial feature of Theorem 4 is the strength of the uniformity norm, which can get arbitrarily close to that of the density norm. This is possible because the Schatten q -norms essentially interpolate between a measure of density (when $q \leq 2$) and a measure of uniformity (when $q = \infty$). Naturally, the bound for the entropies $\mathbf{H}(\mathcal{P}), \mathbf{H}(\mathcal{Q})$ blows up as r approaches q .

Remark 6. The case $q = 2$ of Theorem 4 is a dense regularity lemma (generalizing Theorem 2), applicable when $\|f\|_{L^2} = O(1)$. Surprisingly, Theorem 4 is also interesting in the “super-dense” regime $q < r = 2$, where the norm $\|f - \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q})\|_{L^2}$ is shown to be small; see Theorem 21.

We give a generalization of Theorem 4, which approximates more functions simultaneously, in Theorem 47; we regard these sparse regularity lemmas and the sparse graph removal lemmas that they lead to (discussed in the next subsection) as our main results. Indeed, equipped with this regularity lemma, a compatible *two-sided* counting lemma (Lemma 51) will follow from standard properties of the Schatten norms. Our transference step will use a *dense pairs condition* similar to, but more general than that of Conlon, Fox, Sudakov and Zhao from [14]. For paths and even-length cycles, we can do better by entirely avoiding the transference step, and using properties related to Sidorenko’s conjecture [43]; Figure 1 outlines the logical dependencies between our main results.

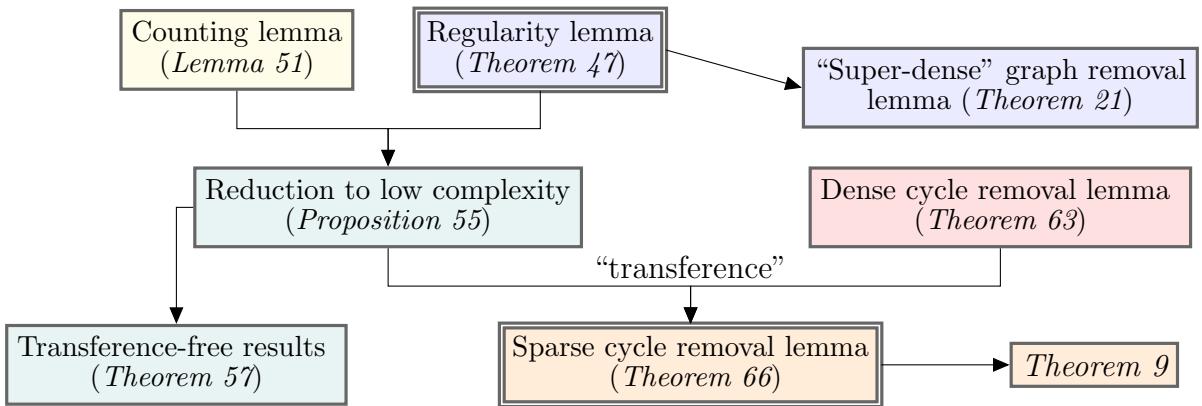


Figure 1: Relationships between main results (arrows show logical implications).

To better understand the regime of applicability of our work, let us mention a combinatorial characterization of $\|f\|_{S^q}$ when $q = 2k$ is an even positive integer: using indices modulo k , one has

$$\|f\|_{S^{2k}}^{2k} = \mathbf{E}_{\substack{x_1, \dots, x_k \\ y_1, \dots, y_k}} \left[\prod_{i=1}^k f(x_i, y_i) \overline{f(x_{i+1}, y_i)} \right], \quad (3)$$

which generalizes (2), and can be viewed as a normalized count of homomorphic copies of $2k$ -cycles in the weighted bipartite graph induced by f (this is closely related to the *trace method* from spectral graph theory [20, 12]). Furthermore, due to the characterization of spectral norms, one has

$$\|f\|_{S^\infty} = \sup_{\|g\|_{L^2} = \|h\|_{L^2} = 1} |\mathbf{E}_{x,y} [f(x, y) g(x) h(y)]|, \quad (4)$$

where g and h range over the complex functions on X , respectively Y , with L^2 norms equal to 1; note the resemblance to the cut norm in (1). In particular,

- (i). $\|f\|_{S^2}$ is the L^2 norm of f ;
- (ii). $\|f\|_{S^4}$ is the \square^2 Gowers box norm of f ;
- (iii). $\|f\|_{S^6}$ is the cube root of the L^2 norm of the *dual* function of f (see, e.g., [51, 48]);
- (iv). $\|f\|_{S^\infty}$, which we may call the *spectral norm* of f , is at least equal to the cut norm $\|f\|_\square$.

Moreover, as they are the L^q norms of the same sequence, the Schatten q -norms of a given function are nonincreasing in q and obey interpolation bounds. Crucially, for $q > 2$, the norm $\|\mathbb{1}_G/p\|_{S^q}$ often remains bounded as $p = \mathbf{E}[\mathbb{1}_G] \searrow 0$, which makes Theorem 4 applicable to sparse graphs.

1.1 Results about graphs

To demonstrate the applicability of our Schatten-norm-based regularity method (including the regularity lemma in Theorem 47 and the counting lemma in Lemma 51), we use it to derive various new graph removal lemmas, particularly in sparse settings.

Remark 7. Lemma 51 also partially answers a question of Conlon, Fox, Sudakov and Zhao [15] about two-sided H -counting lemmas in 4-cycle-free graphs, when H is a cycle or a path; see Corollary 53 and the remark preceding it. In fact, Corollary 53 applies to any graph with $O(n^2)$ 4-cycles, but it requires a slightly stronger uniformity assumption than in [15].

Unless specified otherwise, all graphs mentioned in this paper are finite, unweighted, undirected, and simple. We denote by $V(G)$ and $E(G)$ the sets of vertices and edges of a graph G ; thus the elements of $E(G)$ are two-element subsets of $V(G)$. We define $\mathbb{1}_G : V(G) \times V(G) \rightarrow \{0, 1\}$ by setting $\mathbb{1}_G(x, y) = 1$ iff $\{x, y\} \in E(G)$, and we write $e_G(V, W) := \sum_{x \in V, y \in W} \mathbb{1}_G(x, y)$ for any $V, W \subset V(G)$. Here are some particular graphs:

- For $k \geq 3$, we denote by C_k the cycle of length k (with k vertices and k edges).
- For $k \geq 1$, we denote by P_k the path of length k (with $k+1$ vertices and k edges).
- For $s, t \geq 1$, we denote by $K_{s,t}$ the complete bipartite graph on $s+t$ vertices (e.g., $K_{2,2} \cong C_4$).

We also distinguish between several meanings for counting instances of a graph in another graph:

- (1). By the number of *copies* of a graph H in a graph G (or simply the number of H 's in G), we mean the number of subgraphs of G which are isomorphic to H .
- (2). We write $\text{Inj}(H, G)$ for the number of *injective copies* of H in G , i.e. the number of monomorphisms from H to G (or embeddings of H in G).
- (3). We write $\text{Hom}(H, G)$ for the number of *homomorphic copies* of H in G , i.e. the number of homomorphisms from H to G .

Formally, we write

$$\begin{aligned} \text{Hom}(H, G) &:= \sum_{x_v \in V(G), \forall v \in V(H)} \prod_{\{u,v\} \in E(H)} \mathbb{1}_G(x_u, x_v), \\ \text{Inj}(H, G) &:= \sum_{x_v \in V(G), \forall v \in V(H)} \mathbb{1}_{x_u \neq x_v, \forall u \neq v} \prod_{\{u,v\} \in E(H)} \mathbb{1}_G(x_u, x_v). \end{aligned}$$

Hence when $V(G)$ is equipped with the uniform probability distribution, (3) implies that

$$\forall k \in \mathbb{Z}_{\geq 2} : \quad \text{Hom}(C_{2k}, G) = n^{2k} \|\mathbb{1}_G\|_{S^{2k}}^{2k}. \quad (5)$$

Remark 8. $\text{Inj}(H, G)$ is the same as the number of copies of H in G up to a factor depending on H ; however, $\text{Inj}(H, G)$ can be much smaller than $\text{Hom}(H, G)$ when G is sparse. Also, in light of (5), a simple consequence of the monotonicity of Schatten q -norms in q is that for a fixed n -vertex graph G , the quantity $\frac{1}{n} \text{Hom}(C_{2k}, G)^{1/2k}$ is nondecreasing in k .

Throughout this subsection, n is understood to be a positive integer, and we write $o(1) = o_{n \rightarrow \infty}(1)$ for a function of n which vanishes as $n \rightarrow \infty$ (see Section 2.1 for other asymptotic notation); thus one can understand our results either as ε - δ statements, or as statements about an implicit sequence of graphs $\{G_n\}_{n \geq 1}$. We will also use a parameter $p > 0$ which may depend on n , and which essentially plays the role of the edge density of G (although we do not explicitly require that $p = \mathbf{E}[\mathbb{1}_G]$); thus for our graphs to be sparse, we need $p = o(1)$.

Our main removal lemma, which uses a transference argument based on a dense pairs condition (and has tower-exponential implied dependencies), is given below; see Theorem 66 for a more general version, applicable to weighted and k -partite graphs.

Theorem 9 (Sparse removal lemma for odd-length cycles). *For any $\varepsilon, K, C > 0$ and integer $k \geq 2$, there exist $n_0, \delta, \tau > 0$ such that the following holds. Let $n \geq n_0$, $p > 0$, and G be an n -vertex graph such that*

$$\text{Hom}(C_{2k}, G) \leq Cp^{2k}n^{2k} \quad \text{and} \quad \text{Inj}(C_{2k+1}, G) \leq \delta p^{2k+1}n^{2k+1}.$$

Suppose that for any partition of $V(G)$ into parts V_1, V_2, \dots with at least τn elements each, a total of at most $\varepsilon pn^2/8$ edges of G lie between ‘dense’ pairs (V_i, V_j) with $i \neq j$ and $e_G(V_i, V_j) \geq Kp|V_i||V_j|$. Then G can be made $\{C_3, C_5, \dots, C_{2k+1}\}$ -free by removing at most εpn^2 edges.

Remark 10. A similar result holds for paths and even-length cycles, except that in the conclusion, the graph G becomes empty after removing εpn^2 edges (essentially because paths and even-length cycles are bipartite). In that case, it makes more sense to state the contrapositive statement (concluding that G contains many injective copies of P_j and C_{2k}), given in Theorem 57. We note that Theorem 57 avoids the transference step (and thus also the tower-exponential dependencies).

Remark 11. Theorem 9 may be compared to [14, Proposition 4.2], which can be recovered by taking $k = 2$ and $p = n^{-1/2}$. The dense pairs condition in [14, Proposition 4.2] is more restrictive, since:

- (1). It applies to all partitions with at most M parts, while our condition only concerns partitions with large parts (and one can always take $M = 1/\tau$);
- (2). It includes the contribution of the diagonal pairs (V_i, V_i) .

Point (1) will be relevant for Corollary 12, while (2) will be relevant for Corollaries 15 and 19.

Corollary 12 (Subgraphs of mildly-pseudorandom graphs). *Fix $k \in \mathbb{Z}_{\geq 2}$ and $K > 0$. Let $p > 0$ and G be an n -vertex graph with $\mathbb{1}_G \leq Kp\nu$, for some $\nu : V(G) \times V(G) \rightarrow [0, \infty)$ satisfying*

$$\|\nu\|_{S^{2k}} = O(1) \quad \text{and} \quad \|\nu - 1\|_{\square} = o(1).$$

If G has $o(p^{2k+1}n^{2k+1})$ C_{2k+1} ’s, it can be made $\{C_3, C_5, \dots, C_{2k+1}\}$ -free by removing $o(pn^2)$ edges.

Remark 13. To make sense of the assumptions about ν in Corollary 12, $V(G)$ is implicitly equipped with uniform probability. Note that one can replace these two assumptions with the single (stronger) condition $\|\nu - 1\|_{S^{2k}} = o(1)$; since $k \geq 2$, this is more general than the standard linear forms conditions [17, 48, 51]. For a better comparison with other notions of pseudorandomness, see the remark after Corollary 71 (which also considers the case of removing triangles, i.e., $2k + 1 = 3$).

Remark 14. One can view $p\nu$ as the indicator function of a weighted graph Γ , of which G is a subgraph. Conlon, Fox and Zhao [16] proved similar H -removal lemmas (for a general graph H) under the assumption that Γ is $(p, o(p^{c(H)}n))$ -jumbled, essentially meaning that

$\|\nu - 1\|_{S^\infty} = o(p^{c(H)-1})$; here $c(H) = 1 + 1/(2k-2)$ when H is a cycle of length $2k+1 \geq 5$. Our pseudorandomness constraint $\|\nu - 1\|_\square = o(1)$ is strictly weaker (both in the type of norm bounded and in the bound itself), though it comes at the expense of the (fairly weak) additional assumption $\|\nu\|_{S^{2k}} = O(1)$.

When $p \geq n^{-1/2}$, our removal lemmas also apply to graphs with few 4-cycles; the corollary below is closely related to the results in [14, 15], although we use a different regularity method.

Corollary 15 (Graphs with not too many 4-cycles). *Fix integers $j \geq 1$, $k \geq 2$, $\ell \geq 3$. For any $p \geq n^{-1/2}$ and any n -vertex graph G with $O(p^4 n^4)$ C_4 's, the following hold true.*

- (i). *If G has $o(p^j n^{j+1})$ P_j 's, then it has $o(pn^2)$ edges.*
- (ii). *If G has $o(p^{2\ell} n^{2\ell})$ $C_{2\ell}$'s, then it has $o(pn^2)$ edges.*
- (iii). *If G has $o(p^{2k+1} n^{2k+1})$ C_{2k+1} 's and every bipartite subgraph of G has at most $o(p^4 n^4)$ C_4 's, then G can be made $\{C_3, C_5, \dots, C_{2k+1}\}$ -free by removing $o(pn^2)$ edges.*

(iii'). If G has $o(p^4 n^4)$ C_4 's and $o(p^{2k+1} n^{2k+1})$ C_{2k+1} 's, it can be made $\{C_3, C_5, \dots, C_{2k+1}\}$ -free by removing $o(pn^2)$ edges.

Remark 16. Parts (i) and (ii) of Corollary 15 follow from Theorem 57, while part (iii) follows from Theorem 9; the latter generalizes the fact that

(iii'). *If G has $o(p^4 n^4)$ C_4 's and $o(p^{2k+1} n^{2k+1})$ C_{2k+1} 's, it can be made $\{C_3, C_5, \dots, C_{2k+1}\}$ -free by removing $o(pn^2)$ edges.*

When $2k+1 = 5$, (iii') recovers the main result of Conlon, Fox, Sudakov and Zhao's first aforementioned paper (see [14, Theorem 1.2]). For a general k , (iii') essentially follows from the same authors' follow-up work on counting lemmas in C_4 -free graphs. Another generalization of (iii'), which applies to k -partite graphs (as required in arithmetic applications), is given in Corollary 76.

In particular, Corollary 15 applies to C_4 -free graphs. More generally, we give an application of Theorem 9 to graphs lacking a *quasi-smooth* family of bipartite graphs, following [32, 2].

Definition 17 (Extremal numbers and quasi-smooth families). Given positive integers m, n and a family of bipartite graphs \mathcal{F} , we denote by $\text{ex}(n, \mathcal{F})$ (respectively, $\text{ex}(m, n, \mathcal{F})$) the maximal number of edges in a graph on n vertices (respectively, a bipartite graph on $m+n$ vertices) which is \mathcal{F} -free, meaning that it contains no (injective) copies of any graph in \mathcal{F} . We then say that \mathcal{F} is (α, β) -quasi-smooth for some reals $2 > \alpha > \beta \geq 1$ iff for $n \geq m \geq 1$, one has

$$\text{ex}(m, n, \mathcal{F}) = O(mn^{\alpha-1} + n^\beta).$$

Remark 18. We do not explicitly require that $\text{ex}(n, \mathcal{F}) \geq cn^\alpha$ for some constant $c > 0$ (as in [32]), although our results will only be nontrivial in that case. With this notation, given integers $t \geq s \geq 2$, Füredi [26] showed that $\{K_{s,t}\}$ is (α, β) -quasi-smooth with $\alpha = 2 - \frac{1}{s}$ and $\beta = 2 - \frac{2}{s}$; we only know the corresponding lower bound $\text{ex}(n, \{K_{s,t}\}) \geq c_{s,t} n^\alpha$ when t is large enough in terms of s [3].

Corollary 19 (Graphs free of a quasi-smooth family). *Fix an integer $k \geq 2$, reals $2 > \alpha > \beta \geq 1$, and let \mathcal{F} be an (α, β) -quasi-smooth family. Then an \mathcal{F} -free n -vertex graph G with*

$$\text{Hom}(C_{2k}, G) = O(n^{2k(\alpha-1)}) \quad \text{and} \quad \text{Inj}(C_{2k+1}, G) = o(n^{(2k+1)(\alpha-1)})$$

can be made $\{C_3, C_5, \dots, C_{2k+1}\}$ -free by removing $o(n^\alpha)$ edges.

Remark 20. As an extension of their work on C_4 -free graphs, Conlon, Fox, Sudakov and Zhao asked [15] about similar removal lemmas in H -free graphs for various bipartite graphs H , including C_{2k} and $K_{s,t}$ for $t \geq s > 2$. Corollary 19 gives a partial answer to this question in the latter case, taking $\mathcal{F} = \{K_{s,t}\}$, $\alpha = 2 - 1/s$ and $\beta = 2 - 2/s$; we note that when $\mathcal{F} = \{K_{s,t}\}$, a generalization of the counting lemmas in [15] should eliminate the need for an upper bound on $\text{Hom}(C_{2k}, G)$. The case $\mathcal{F} = \{C_{2k}\}$ remains open, since this \mathcal{F} is not known to be quasi-smooth for any $k \geq 3$ [32, 2].

Separately from our results for sparse graphs, we also state a natural consequence of our regularity lemma in the regime where $q < r = 2$; this is a graph removal lemma for a family of “super-dense” graphs (with bounded Schatten q -norm for some $q < 2$), involving no tower-exponential bounds. Its proof is left to Appendix B.

Theorem 21 (Super-dense graph removal lemma). *For any graph H , $q \in [1, 2)$, $C > 0$ and $\varepsilon \in (0, 1)$, there exists*

$$\delta \geq \exp(-O_{H,q,C}(\varepsilon^{-O_{H,q}(1)}))$$

such that the following holds. Let G be an n -vertex graph satisfying $\|\lambda^G\|_q \leq Cn$, where $(\lambda_i^G)_{i \geq 1}$ are the eigenvalues of G ’s adjacency matrix. If $\text{Hom}(H, G) \leq \delta n^{|V(H)|}$, then one can remove at most εn^2 edges from G to eliminate all homomorphic copies of H .

Remark 22. As shown in [37], most graphs G have $\|\lambda^G\|_q$ of order $n^{\frac{1}{2} + \frac{1}{q}}$ when $q \in [1, 2)$, which is not $\leq Cn$; however, Theorem 21 should be interesting when $q = 2 - \varepsilon$, since one always has $\|\lambda^G\|_2 \leq n$.

1.2 Corollaries in additive combinatorics

We now give examples of applications for each of Corollaries 12, 15 and 19 (more precisely, of their k -partite generalizations), to finding translation-invariant patterns in certain sparse subsets of abelian groups. This makes use of bipartite graphs with weights of the form $f(x, y) := \varphi(x + y)$, for some function $\varphi : G \rightarrow [0, \infty)$, where G is an abelian group of order n . Equipping $G \times G$ with the uniform probability measure, we extend the meaning of the cut and Schatten norms by setting

$$\|\varphi\|_\square := \|f\|_\square \quad \text{and} \quad \|\varphi\|_{S^q} := \|f\|_{S^q}, \quad \forall q \in [1, \infty].$$

In this context, $\|\varphi\|_{S^q}$ is a normalized L^q norm on the Fourier coefficients of φ , due to Lemma 40. When $q = 2k$ is an even positive integer, we also have

$$n^{2k-1} \|\varphi\|_{S^{2k}}^{2k} = \sum_{a_1 + \dots + a_k = b_1 + \dots + b_k} \prod_{i=1}^k \varphi(a_i) \overline{\varphi(b_i)},$$

which is often referred to as *k-fold additive energy* (especially when φ is an indicator function).

We first give an application in the setting of a pseudorandom majorant. As in Corollary 12, the key feature is that the pseudorandomness condition is mild and easy to verify, compared to the linear forms conditions in [17, 48, 51] and the jumbledness requirements in [16]. In particular, one can apply the following result to find solutions to linear equations in the primes (see also Corollary 74 and the remark after Corollary 71 for the case $k = 3$).

Corollary 23 (Linear patterns in subsets of mildly-pseudorandom sets). *For any $\varepsilon, C > 0$ and $k, \ell \in \mathbb{Z}$ with $k > 2\ell \geq 4$, there exists $\delta > 0$ such that the following holds. Let a_1, \dots, a_k be nonzero integers with $a_1 + \dots + a_k = 0$, G be an abelian group of order n with $\gcd(n, a_1 \cdots a_k) = 1$, and $S \subset G$, $\nu : G \rightarrow [0, \infty)$ satisfy*

- (i) (ν majorizes 1_S after scaling). $\varepsilon 1_S \leq \frac{|S|}{n} \nu$,
- (ii) (ν is mildly-pseudorandom). $\|\nu\|_{S^{2\ell}} \leq C$ and $\|\nu - 1\|_{\square} \leq \delta$.

Then S contains $\geq \delta |S|^k / n$ solutions in x_1, \dots, x_k to the linear equation $a_1 x_1 + \dots + a_k x_k = 0$.

Remark 24. Choosing $C > 2$ and $\delta < 1$, condition (ii) follows from the simpler constraint $\|\nu - 1\|_{S^{2\ell}} \leq \delta$.

Remark 25. The equation $a_1 x_1 + \dots + a_k x_k = 0$ has $|S|$ trivial solutions $x_1 = \dots = x_k \in S$, so Corollary 23 only has nontrivial content when $|S|^{k-1} > \delta^{-1} n$. Also, note that a random subset of G with $|S|$ elements is expected to contain $|S|^k / n$ solutions to such equations.

Before we state our other applications, let us recall that a subset S of an abelian group is called *Sidon* if and only if it contains no nontrivial solutions to the equation $x + t = y + z$ (the trivial solutions being $(x, t) = (y, z)$ and $(x, t) = (z, y)$); this is equivalent to the absence of 4-cycles in the associated bipartite graph, where (x, y) is an edge whenever $x + y \in S$. Such sets have $|S| = O(\sqrt{n})$.

Prendiville [38] proved a result similar to Corollary 23, where S is a large “almost-Sidon” subset of $[n] := \{1, \dots, n\}$ (meaning that S contains few nontrivial solutions to $x + t = y + z$, and $|S| \geq c\sqrt{n}$). One can also deduce this result, with weaker implied bounds, from the k -partite version of Corollary 15. Below we consider another natural weakening of the Sidon condition, noting that a set is Sidon iff it contains at most one pair of the form $(x, x + h)$ for each $h \neq 0$.

Corollary 26 (Linear patterns in generalized Sidon sets). *For any $\varepsilon, C > 0$, $k \in \mathbb{Z}_{\geq 5}$, and nonzero integers a_1, \dots, a_k with $a_1 + \dots + a_k = 0$, there exists $\delta > 0$ such that the following holds. Let $n \in \mathbb{Z}_{\geq 1}$ and $S \subset [n]$ satisfy*

- (i) (S is large). $|S| \geq \varepsilon n^{1/2}$, and
- (ii) (Sidon-type condition). For each $h \neq 0$, S contains at most C pairs of the form $(x, x + h)$.

Then S contains $\geq \delta \frac{|S|^k}{n}$ solutions in x_1, \dots, x_k to the linear equation $a_1x_1 + \dots + a_kx_k = 0$.

Remark 27. Corollary 26 follows from (the k -partite version of) Corollary 19 and the fact that $K_{2,t}$ is quasi-smooth for all $t \geq 2$. One obtains a more general statement using a general complete bipartite graph $K_{s,t}$ instead; see Corollary 82.

Finally, as an application of Corollary 15, we give a two-dimensional analogue of Prendiville's result [38], related to the multidimensional Roth theorem [44, 45, 28]. Given slopes $s_1, \dots, s_k \in \mathbb{Q} \cup \{\infty\}$, let us call a sequence of lattice points $P = (p_1, \dots, p_k) \in (\mathbb{Z}^2)^k$ an (s_1, \dots, s_k) -polygon iff for each $i \in \mathbb{Z}/k\mathbb{Z}$, the points p_i and p_{i+1} lie on a line of slope s_i . We say that P is non-degenerate when p_1, \dots, p_k are pairwise distinct.

Corollary 28 (Polygonal patterns in large sets with few parallelograms). *For any $\varepsilon > 0$, $k \in \mathbb{Z}_{\geq 5}$, and slopes $s_i \in \mathbb{Q} \cup \{\infty\}$ with $s_i \neq s_{i+1}$ for $i \in \mathbb{Z}/k\mathbb{Z}$, there exists $\delta > 0$ such that the following holds. Let $n \in \mathbb{Z}_{\geq 1}$ and $S \subset [n] \times [n]$ satisfy*

- (i) (S is large). $|S| \geq \varepsilon n^{3/2}$, and
- (ii) (Few parallelograms). For each $i \in \mathbb{Z}/k\mathbb{Z}$, S contains at most $\delta |S|^4/n^4$ non-degenerate $(s_i, s_{i+1}, s_i, s_{i+1})$ -polygons.

Then S contains at least $\delta \frac{|S|^k}{n^k}$ (possibly degenerate) (s_1, \dots, s_k) -polygons.

Remark 29. It can be shown, using (21), that any set $S \subset [n] \times [n]$ which satisfies condition (ii) of Corollary 28 with small enough δ has $|S| = O(n^{3/2})$; thus Corollary 28 concerns the densest (up to a constant) sets with few parallelograms.

2 Notation and preliminaries

2.1 Asymptotic notation, sets and functions

We keep the notation $[n] := \{1, \dots, n\}$ when n is a positive integer. We may add restrictions in the subscripts of $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ to denote the appropriate subsets; for instance, $\mathbb{Z}_{\geq a}$ denotes the set of all integers greater than or equal to a . Given a set $S \subset \mathbb{C}$ and some $k \in \mathbb{C}$, we also write $kS := \{ks : s \in S\}$; for instance, $2\mathbb{Z}$ denotes the set of all even integers. For $r \in \mathbb{R}$, we write $\lfloor r \rfloor$ for the greatest integer smaller than or equal to r .

Given any functions $f : X \rightarrow \mathbb{C}$ and $g : X \rightarrow [0, \infty)$, we write $f = O(g)$ or $f \ll g$ iff there is an absolute constant C such that $|f| \leq Cg$ pointwise. We write $f \asymp g$ when $f \ll g$ and $g \ll f$. If the implied constants may depend on a parameter q , we indicate this using subscripts: $f = O_q(g)$, $f \ll_q g$, respectively $f \asymp_q g$. If $X = \mathbb{Z}_{\geq N}$ for some $N \geq 1$, we write $f = o(g) = o_{n \rightarrow \infty}(g)$ iff $f(n)/g(n) \rightarrow 0$ as $n \rightarrow \infty$.

Given a subset $S \subset X$ of an implicit ambient space, we write $\mathbb{1}_S : X \rightarrow \{0, 1\}$ for the indicator function of S . When S is an statement such as $x \neq y$, we may write $\mathbb{1}_S$ for its truth value. We also keep all the graph-related notations from Section 1.1; in particular, when G is a graph, we have $\mathbb{1}_G(x, y) = \mathbb{1}_{x \neq y} \mathbb{1}_{\{(x, y) \in E(G)\}}$.

Given a measurable function $f : (X, \Sigma, \mu) \rightarrow \mathbb{C}$ and some $q \in [1, \infty]$, we denote the L^q norm of f by $\|f\|_{L^q}$ when μ is a probability measure, and by $\|f\|_q$ when μ is the counting measure (to avoid potential confusion). We also write $\|v\|_q$ for the L^q norms of vectors $v \in \mathbb{C}^n$.

Finally, we define the Fourier transform of a function $\varphi : G \rightarrow \mathbb{C}$ on a finite abelian group G by

$$\widehat{\varphi}(\chi) := \sum_{g \in G} f(g) \overline{\chi(g)}, \quad \forall \chi \in \widehat{G}. \quad (6)$$

2.2 Probability spaces, partitions and entropy

We will mostly work with finite probability spaces (X, Σ, \mathbf{P}) . If the σ -algebra Σ is not specified, it should be understood that $\Sigma = 2^X$ (so \mathbf{P} can be identified with a function from X to $[0, 1]$ with total sum 1). If the probability measure \mathbf{P} is not specified, it should be understood that \mathbf{P} is the uniform probability measure on X (especially when $X = V(G)$ for some graph G). Given an event $E \in \Sigma$, we denote its complement by $E^c := X \setminus E$. Given an absolutely integrable function $f : X \rightarrow \mathbb{C}$, we write $\mathbf{E}[f]$ for the expected value of f , and we may omit the square brackets when the content of the expectation is unambiguous. Given an event $A \in \Sigma$ with positive probability, we also write

$$\mathbf{E}[f \mid A] := \frac{\mathbf{E}[f \mathbb{1}_A]}{\mathbf{E}[\mathbb{1}_A]}$$

for the conditional expected value of f given A . If $X = X_1 \times \dots \times X_k$ where $(X_i, \Sigma_i, \mathbf{P}_i)$ are probability spaces, then in an expression like $\mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k)$, it should be understood that the x_i 's are sampled independently from (X_i, \mathbf{P}_i) (thus $\mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) = \mathbf{E}[f]$, using the implicit product probability measure on $X_1 \times \dots \times X_k$).

A *partition* \mathcal{P} of a set X is a collection of pairwise disjoint subsets of X (called *parts*) with union equal to X . Hence $|\mathcal{P}|$ denotes the cardinality of this collection, i.e., the number of parts. We also denote $\text{diag}(\mathcal{P}) := \bigcup_{A \in \mathcal{P}} (A \times A)$, which is a subset of $X \times X$.

If (X, Σ) is a measurable space, we will only work with measurable partitions of X (meaning that all parts belong to Σ); conversely, we write $\Sigma(\mathcal{P})$ for the smallest σ -algebra containing all the parts of a partition \mathcal{P} . Given a measurable function $f : (X, \Sigma) \rightarrow S$, we write $\text{Part}(f)$ for the partition of X induced by f , i.e., $\text{Part}(f) := \{f^{-1}(s) : s \in S\}$. Given an absolutely integrable function $f : (X, \Sigma, \mathbf{P}) \rightarrow S$, we write $\mathbf{E}(f \mid \mathcal{P}) := \mathbf{E}(f \mid \Sigma(\mathcal{P}))$ for the conditional expectation of f with respect to \mathcal{P} ; if X is finite, we can view $\mathbf{E}(f \mid \mathcal{P})$ as an averaged function over parts of \mathcal{P} :

$$\mathbf{E}(f \mid \mathcal{P})(x) = \mathbf{E}[f \mid P_x], \quad \text{where } x \in P_x \in \mathcal{P}. \quad (7)$$

Given two partitions \mathcal{P}, \mathcal{Q} of a set X , we say that \mathcal{P} is finer than \mathcal{Q} iff $\Sigma(\mathcal{P})$ is richer than $\Sigma(\mathcal{Q})$ (i.e., $\Sigma(\mathcal{P})$ contains all parts of \mathcal{Q}). We write $\mathcal{P} \wedge \mathcal{Q}$ for the common refinement of \mathcal{P} and \mathcal{Q} , i.e., the coarsest partition which is finer than or equal to both \mathcal{P} and \mathcal{Q} . Given partitions \mathcal{P} of X and \mathcal{Q} of Y , we write $\mathcal{P} \otimes \mathcal{Q}$ for the product partition $\{A \times B : A \in \mathcal{P}, B \in \mathcal{Q}\}$ of $X \times Y$.

Finally, given a probability space (X, Σ, \mathbf{P}) and a measurable partition \mathcal{P} of X with finitely many parts, we define its (Shannon) *entropy* by

$$\mathbf{H}(\mathcal{P}) := \sum_{A \in \mathcal{P}} \mathbf{P}(A) \log \frac{1}{\mathbf{P}(A)},$$

which increases when we pass to a finer partition; thus we have $0 = \mathbf{H}(\{X\}) \leq \mathbf{H}(\mathcal{P}) \leq \mathbf{H}(\{\{x\} : x \in X\}) = \log |X|$. Given a finitely-valued random variable $f : X \rightarrow S$, we also define $\mathbf{H}(f) := \mathbf{H}(\text{Part}(f)) = -\sum_{s \in S} \mathbf{P}(f = s) \log \mathbf{P}(f = s)$ (this is the most common definition of entropy). We will only use the inequality

$$\mathbf{H}(\mathcal{P} \wedge \mathcal{Q}) \leq \mathbf{H}(\mathcal{P}) + \mathbf{H}(\mathcal{Q}), \quad (8)$$

which translates to $\mathbf{H}(f, g) \leq \mathbf{H}(f) + \mathbf{H}(g)$ in random-variable notation; this follows from Jensen's inequality, and the difference $\mathbf{I}(\mathcal{P} : \mathcal{Q}) := \mathbf{H}(\mathcal{P}) + \mathbf{H}(\mathcal{Q}) - \mathbf{H}(\mathcal{P} \wedge \mathcal{Q})$ is called the *mutual information* between \mathcal{P} and \mathcal{Q} . We note that this property is also true for the simpler function $\log |\mathcal{P}|$, but entropy is less sensitive to parts with small probability (which is crucial to our regularity lemma).

We write M^T and M^* for the transpose and the conjugate transpose of a matrix M . We now introduce a couple of matrices related to random variables and partitions, all in finite probability spaces with full σ -algebras.

Definition 30 (Normalized matrices). Given finite probability spaces (X, \mathbf{P}_X) , (Y, \mathbf{P}_Y) , a function $f : X \times Y \rightarrow \mathbb{C}$, and an ordering $(x_i)_{1 \leq i \leq |X|}$ and $(y_j)_{1 \leq j \leq |Y|}$ of the elements of X and Y , we define $\text{Mat}(f)$ as the $|X| \times |Y|$ complex matrix with entries

$$\text{Mat}(f)_{i,j} := f(x_i, y_j) \sqrt{\mathbf{P}_X(x_i) \mathbf{P}_Y(y_j)}.$$

Remark 31. To motivate this normalization, one can check that given any $k \geq 2$, finite probability spaces (X_i, \mathbf{P}_i) (with orderings of their elements), and functions $f_i : X_i \rightarrow \mathbb{C}$ for $i \in \mathbb{Z}/k\mathbb{Z}$, one has

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) = \text{Tr}(\text{Mat}(f_1) \cdots \text{Mat}(f_k)). \quad (9)$$

Definition 32 (Averaging matrices). Let (X, \mathbf{P}_X) be a finite probability space, and let $(x_i)_{1 \leq i \leq |X|}$ be an ordering of the elements of X . Then for any partition \mathcal{P} of X , we define the matrix $A_{\mathcal{P}} \in [0, \infty)^{n \times n}$ (which also implicitly depends on \mathbf{P}_X and on the chosen ordering $(x_i)_i$) by

$$(A_{\mathcal{P}})_{i,j} := \begin{cases} \frac{\sqrt{\mathbf{P}_X(x_i) \mathbf{P}_X(x_j)}}{\mathbf{P}_X(P)}, & \exists P \in \mathcal{P} : x_i, x_j \in P, \\ 0, & \text{otherwise.} \end{cases}$$

Hence $A_{\mathcal{P}}$ is a block-diagonal matrix with $|\mathcal{P}|$ diagonal blocks, the sum of the squared entries in each block being 1, and each block having rank 1 (in particular, $\text{rank}(A_{\mathcal{P}}) = |\mathcal{P}|$).

Lemma 33 (Properties of averaging matrices). *Fix finite probability spaces (X, \mathbf{P}_X) and (Y, \mathbf{P}_Y) , as well as orderings of their elements. With the understanding that \mathcal{P} and \mathcal{R} are partitions of X and \mathcal{Q} is a partition of Y , the following hold true.*

(i). *For any $f : X \times Y \rightarrow \mathbb{C}$, one has*

$$\text{Mat}(\mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q})) = A_{\mathcal{P}} \text{Mat}(f) A_{\mathcal{Q}}.$$

(ii). *If \mathcal{R} is finer than or equal to \mathcal{P} , then*

$$A_{\mathcal{P}} A_{\mathcal{R}} = A_{\mathcal{R}} A_{\mathcal{P}} = A_{\mathcal{P}}.$$

In particular, one has $A_{\mathcal{P}}^2 = A_{\mathcal{P}}$, so $A_{\mathcal{P}}$ is a projection. Another consequence of this is that $(A_{\mathcal{R}} - A_{\mathcal{P}})^2 = A_{\mathcal{R}} - A_{\mathcal{P}}$, so $A_{\mathcal{R}} - A_{\mathcal{P}}$ is also a projection.

(iii). *For any complex matrix M with $|X|$ rows, denoting $\|A\|_{S^2} := \text{Tr}(AA^*)^{1/2}$, one has*

$$\|(A_{\mathcal{R}} - A_{\mathcal{P}})M\|_{S^2}^2 = \|A_{\mathcal{R}}M\|_{S^2}^2 - \|A_{\mathcal{P}}M\|_{S^2}^2. \quad (10)$$

(iv). *The nonzero singular values (and eigenvalues) of $A_{\mathcal{P}}$ consist of $|\mathcal{P}|$ ones. If \mathcal{R} is finer than \mathcal{P} , then the nonzero singular values (and eigenvalues) of $A_{\mathcal{R}} - A_{\mathcal{P}}$ consist of $|\mathcal{R}| - |\mathcal{P}|$ ones.*

Proof. Parts (i) and (ii) follow quickly from our definitions and (7). Part (iii) is Pythagoras' theorem; more precisely, we can expand

$$\begin{aligned} \text{Tr}((A_{\mathcal{R}} - A_{\mathcal{P}})MM^*(A_{\mathcal{R}} - A_{\mathcal{P}})) &= \text{Tr}((A_{\mathcal{R}} - A_{\mathcal{P}})^2MM^*) \\ &= \text{Tr}((A_{\mathcal{R}} - A_{\mathcal{P}})MM^*) \\ &= \text{Tr}(A_{\mathcal{R}}MM^*) - \text{Tr}(A_{\mathcal{P}}MM^*) \\ &= \text{Tr}(A_{\mathcal{R}}MM^*A_{\mathcal{R}}) - \text{Tr}(A_{\mathcal{P}}MM^*A_{\mathcal{P}}), \end{aligned}$$

using the circular symmetry and the linearity of the trace. Part (iv) follows from the fact that projections can only have eigenvalues of 0 and 1, and from (10) when M is the identity matrix (thus $A_{\mathcal{R}} - A_{\mathcal{P}}$ has a total of $\|A_{\mathcal{R}}\|_{S^2}^2 - \|A_{\mathcal{P}}\|_{S^2}^2 = |\mathcal{R}| - |\mathcal{P}|$ singular values equal to 1). \square

2.3 Schatten norms and related energies

Definition 34 (Schatten norms of matrices). Given $q \in [1, \infty)$ and a matrix $M \in \mathbb{C}^{m \times n}$ with singular values $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min(m,n)}$, we define its *Schatten norms* by

$$\|M\|_{S^q} := \left(\sum_{j=1}^{\min(m,n)} \sigma_j^q \right)^{1/q} \quad \text{and} \quad \|M\|_{S^\infty} := \max_{1 \leq j \leq \min(m,n)} \sigma_j.$$

Lemma 35. For $q, q_i, r \in [1, \infty]$ and any suitably-sized complex matrices M, A_1, \dots, A_k , the following hold true.

(i). (Norm) $\|\cdot\|_{S^q}$ is a unitary-invariant norm on $\mathbb{C}^{m \times n}$, and $\|M\|_{S^q} = \|M^T\|_{S^q} = \|M^*\|_{S^q}$.

(ii). (Monotonicity and interpolation) For $q_0 < q_1$, one has $\|M\|_{S^{q_0}} \geq \|M\|_{S^{q_1}}$. Moreover, for $\theta \in [0, 1]$ and $\frac{1}{q_\theta} := \frac{1-\theta}{q_0} + \frac{\theta}{q_1}$, one has

$$\|M\|_{S^{q_\theta}} \leq \|M\|_{S^{q_0}}^{1-\theta} \cdot \|M\|_{S^{q_1}}^\theta.$$

(iii). (Special cases) $\|M\|_{S^\infty}$ is the spectral (L^2 operator) norm of M , often written $\|M\|$ or $\|M\|_2$, while $\|M\|_{S^2}$ is the L^2 norm on the entries of M (also known as the Frobenius norm). More generally, one has $\|M\|_{S^{2k}}^{2k} = \text{Tr}((MM^*)^k)$ for $k \in \mathbb{Z}_{\geq 1}$. In particular, if M is the adjacency matrix of an undirected graph G and $k \geq 2$, then $\|M\|_{S^{2k}}^{2k} = \text{Hom}(C_{2k}, G)$.

(iv). (Doubling property) One has

$$\|M\|_{S^q} = \|MM^*\|_{S^{q/2}}^{1/2}. \quad (11)$$

(v). (Hölder inequality) If $\frac{1}{r} = \frac{1}{q_1} + \dots + \frac{1}{q_k}$, then $\|A_1 \dots A_k\|_{S^r} \leq \|A_1\|_{S^{q_1}} \dots \|A_k\|_{S^{q_k}}$. In particular, given any diagonal matrix I_A with entries in $\{0, 1\}$, and any partitions \mathcal{P}, \mathcal{R} of a suitably-sized set such that \mathcal{R} is finer than or equal to \mathcal{P} , one has

$$\max(\|I_A M\|_{S^r}, \|A_{\mathcal{P}} M\|_{S^r}, \|(A_{\mathcal{R}} - A_{\mathcal{P}})M\|_{S^r}) \leq \|M\|_{S^r}. \quad (12)$$

(vi). (Trace norm inequality) When $q = 1$, $\|M\|_{S^1}$ (also known as the trace norm or nuclear norm) satisfies

$$|\text{Tr}(M)| \leq \|M\|_{S^1}.$$

Proof. For the fact that $\|\cdot\|_{S^q}$ defines a norm, see [6, 53]. Otherwise, parts (i), (iii) and (iv) follow from basic properties of singular values, and part (ii) follows from the monotonicity and interpolation properties of L^q norms (for the counting measure).

The Hölder inequality in (v) reduces to the case $k = 2$ (i.e., $\|AB\|_{S^r} \leq \|A\|_{S^p} \|B\|_{S^q}$ when $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$); we refer the reader to [53, Corollary 4.27] and [4, §III] for this case. Then, (12) follows by taking $p = \infty$ and using that $\|I_A\|_{S^\infty}, \|A_{\mathcal{P}}\|_{S^\infty}, \|A_{\mathcal{R}} - A_{\mathcal{P}}\|_{S^\infty}$ are either 0 or 1 (by Lemma 33).

Finally, let us give a short proof of the trace norm inequality in (vi). Let $M = U\Sigma V^*$ be a singular value decomposition of M , where Σ has diagonal entries σ_i . Then

$$\text{Tr}(M) = \text{Tr}(U\Sigma V^*) = \text{Tr}(\Sigma V^* U) = \sum_i \sigma_i w_i,$$

where w_i is the (i, i) th entry of V^*U , i.e., the inner product between U 's i th column and V 's i th row. But $|w_i| \leq 1$ by Cauchy–Schwarz (since U, V are unitary), and thus $|\text{Tr}(M)| \leq \sum_i \sigma_i = \|M\|_{S^1}$. \square

Definition 36 (Schatten norms of functions). Given $q \in [1, \infty]$, finite probability spaces (X, \mathbf{P}_X) , (Y, \mathbf{P}_Y) , and a function $f : X \times Y \rightarrow \mathbb{C}$, we define the q -Schatten norm of f by

$$\|f\|_{S^q} := \|\text{Mat}(f)\|_{S^q},$$

Note that this does not depend on the chosen ordering of the elements of X and Y (implicit in the notation $\text{Mat}(f)$), since permutation matrices are unitary.

Lemma 37. *Given finite probability spaces (X, \mathbf{P}_X) , (Y, \mathbf{P}_Y) , a function $f : X \times Y \rightarrow \mathbb{C}$, and $q, q_i \in [1, \infty]$, the following hold true.*

- (i). $\|\cdot\|_{S^q}$ defines a norm on functions from $X \times Y$ to \mathbb{C} , which satisfies the same monotonicity and interpolation properties as in Lemma 35. Moreover, $\|f^T\|_{S^q} = \|f\|_{S^q}$.
- (ii). When $q \in 2\mathbb{Z} \cup \{\infty\}$, one can rewrite $\|f\|_{S^q}$ as in (3) and (4). In particular, one has $\|f\|_{S^2} = \|f\|_{L^2}$, $\|f\|_{S^4} = \|f\|_{\square^2}$, and $\|f\|_{S^\infty} \geq \|f\|_\square$.
- (iii). For any $A \subset X$ and $B \subset Y$, and any partitions \mathcal{P} of X and \mathcal{Q} of Y , one has

$$\max(\|f \mathbb{1}_{A \times B}\|_{S^q}, \|\mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q})\|_{S^q}) \leq \|f\|_{S^q}.$$

Moreover, $\text{Mat}(\mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q}))$ has at most $\max(|\mathcal{P}|, |\mathcal{Q}|)$ nonzero singular values; in particular, the constant function $\mathbb{1}_{X \times Y}$ has $\|\mathbb{1}_{X \times Y}\|_{S^q} = 1$.

- (iv). (Cycle counting inequality) Let $k \in \mathbb{Z}_{\geq 2}$ and suppose that $\sum_{i=1}^k \frac{1}{q_i} \geq 1$. For $i \in \mathbb{Z}/k\mathbb{Z}$, let (X_i, \mathbf{P}_i) be finite probability spaces and $f_i : X_i \times X_{i+1} \rightarrow \mathbb{C}$. Then one has

$$\left| \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) \right| \leq \prod_{i=1}^k \|f_i\|_{S^{q_i}}.$$

Proof. Parts (i) and (ii) follow from the first three parts of Lemma 35, (9), and the fact that the operator norm of an $m \times n$ complex matrix M can be rewritten as

$$\|M\|_{S^\infty} = \sup_{\substack{v \in \mathbb{C}^m, w \in \mathbb{C}^n \\ \|v\|_2 = \|w\|_2 = 1}} |v^T M w|,$$

where we can make the change of variables $v_j := g(x_j) \sqrt{\mathbf{P}(x_j)}$ and $w_j := h(y_j) \sqrt{\mathbf{P}(y_j)}$ for some functions $g : X \rightarrow \mathbb{C}, h : Y \rightarrow \mathbb{C}$ with L^2 norms equal to 1. The first part of (iii) follows from Lemma 35.(v), and the second part follows from the facts that $\text{Mat}(\mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{Q})) = A_{\mathcal{P}} \text{Mat}(f) A_{\mathcal{Q}}$, $\text{rank}(A_{\mathcal{P}}) = |\mathcal{P}|$, and $\text{rank}(A_{\mathcal{Q}}) = |\mathcal{Q}|$.

Finally, let us prove the cycle counting inequality in (iv). Fixing an ordering of the elements of each X_i and letting $F_i := \text{Mat}(f_i)$, we need to show (by (9)) that

$$|\text{Tr}(F_1 \cdots F_k)| \leq \prod_{i=1}^k \|F_i\|_{S^{q_i}}.$$

For $i \in \mathbb{Z}/k\mathbb{Z}$, define $q'_i \in [q_i, \infty]$ by $q'_i := q_i \sum_{j=1}^k \frac{1}{d_j}$, so that $\sum_{i=1}^k \frac{1}{q'_i} = 1$. Then by the trace norm and Hölder-type inequalities in Lemma 35, and the monotonicity of Schatten norms, we have

$$|\mathrm{Tr}(F_1 \cdots F_k)| \leq \|F_1 \cdots F_k\|_{S^1} \leq \prod_{i=1}^k \|F_i\|_{S^{q'_i}} \leq \prod_{i=1}^k \|F_i\|_{S^{q_i}},$$

as we wanted. \square

Remark 38. The cycle counting inequality in Lemma 37 can be regarded as a generalization of the Gowers–Cauchy–Schwarz inequality of order 2 (the latter is the special case $k = 4$, $q_1 = \cdots = q_4 = 1/4$). When $q_1 = \cdots = q_k \in 2\mathbb{Z}$, it also follows from Hatami’s Hölder-type inequality [31].

In fact, when q, q_1, \dots, q_k are even integers, most (if not all) of the properties in Lemmas 35 and 37 can be proven combinatorially, using repeated (and rather convoluted) applications of Cauchy–Schwarz, similar to those in [19]. Following this idea, it is possible that some of the generalized Gowers norms considered by Hatami and then Conlon and Lee [31, 19] could function as a higher-order substitute for the Schatten norms, in an extension of our work to hypergraphs.

Open Problem 39. Obtain compatible versions of Theorem 47 and Lemma 51 for hypergraphs, using combinatorial analogues of the Schatten norms.

We also mention a relationship between Schatten norms and L^q norms of Fourier coefficients, connecting the graph-theoretic and Fourier-analytic methods in additive combinatorics.

Lemma 40 (Fourier coefficients and singular values). *Let G be a finite abelian group (equipped with the uniform probability measure) and $a, b \in \mathbb{Z}$ be such that $\gcd(|G|, ab) = 1$. Let $f : G \rightarrow \mathbb{C}$, and define $\tilde{f} : G \times G \rightarrow \mathbb{C}$ by $\tilde{f}(x, y) := f(ax + by)$. Then one has*

$$\|\tilde{f}\|_{S^q} = \frac{1}{|G|} \|\hat{f}\|_q.$$

Proof. Since the maps $x \mapsto ax$ and $y \mapsto by$ are automorphisms of G , and permutations of the rows and columns of a matrix preserve its singular values, we have

$$\|\tilde{f}\|_{S^q} = \frac{1}{|G|} \|M\|_{S^q},$$

where M is the matrix with entries $M_{x,y} = f(x + y)$ (for some ordering of the elements of G). It thus suffices to show that the singular values of M are precisely the absolute values of the Fourier coefficients of f . Indeed, the inverse Fourier transform identity implies that for any $\chi \in \widehat{G}$, one has

$$f(x + y) = \frac{1}{|G|} \sum_{\chi} \chi(x) \widehat{f}(\chi) \chi(y),$$

which translates to the identity of $|G| \times |G|$ matrices (noting that $|G| = |\widehat{G}|$)

$$M = UDU^T, \quad \text{where:} \quad U_{x,\chi} = \chi(x), \quad D_{\chi_1,\chi_2} = \widehat{f}(\chi_1) \mathbb{1}_{\chi_1=\chi_2}.$$

Moreover, the matrices U and U^T are unitary by standard properties of the Fourier transform. Hence $MM^* = UDD^*U^*$, which is a diagonalization of MM^* , and thus the singular values of M are the square roots of the diagonal entries of DD^* , i.e., $|\widehat{f}(\chi)|$ for $\chi \in \widehat{G}$. \square

We end this subsection by defining a family of energy functions related to the Schatten norms, which will be used in the proof of our regularity lemma (and may be of independent interest).

Definition 41 (Generalized L^2 energy). Fix a finite probability space (X, \mathbf{P}_X) and an ordering of its elements. For any complex matrix F with $|X|$ rows, any integer $\ell \geq 1$, and any partition \mathcal{P} of X , we define the (left) ℓ th L^2 energy of \mathcal{P} with respect to F by

$$\mathcal{E}_{F,\ell}(\mathcal{P}) := \|A_{\mathcal{P}}F_{\ell}\|_{S^2}^2,$$

where F_{ℓ} can be defined inductively by $F_1 := F$ and $F_{\ell+1} := F_{\ell}F_{\ell}^*$. In other words,

$$F_{\ell} := \begin{cases} F, & \ell = 1, \\ (FF^*)^{2^{\ell-2}}, & \ell \geq 2. \end{cases}$$

We also define the *total* ℓ th L^2 energy of F by $\mathcal{E}_{F,\ell} := \|F_{\ell}\|_{S^2}^2$. Given another probability space (Y, \mathbf{P}_Y) and a function $f : X \times Y \rightarrow \mathbb{C}$, we define $\mathcal{E}_{f,\ell}(\mathcal{P}) := \mathcal{E}_{\text{Mat}(f),\ell}(\mathcal{P})$ and $\mathcal{E}_{f,\ell} := \mathcal{E}_{\text{Mat}(f),\ell}$.

Remark 42. With the notation above, for $\ell \geq 1$ one has

$$\mathcal{E}_{F,\ell} = \text{Tr}(F_{\ell}F_{\ell}^*) = \text{Tr}(F_{\ell+1}) = \text{Tr}\left((FF^*)^{2^{\ell-1}}\right) = \|F\|_{S^{2^{\ell}}}^{2^{\ell}},$$

and thus $\mathcal{E}_{f,\ell} = \|f\|_{S^{2^{\ell}}}^{2^{\ell}}$ for any $f : X \times Y \rightarrow \mathbb{C}$. Also, $\mathcal{E}_{f,\ell}(\mathcal{P})$ can only increase when we pass to a finer partition, due to (10).

Lemma 43 (Energy increment). *Let (X, \mathbf{P}_X) , (Y, \mathbf{P}_Y) be finite probability spaces, $f : X \times Y \rightarrow \mathbb{C}$, and $\mathcal{P} = \mathcal{P}_X \otimes \mathcal{P}_Y$, $\mathcal{R} = \mathcal{R}_X \otimes \mathcal{R}_Y$ be partitions of $X \times Y$ such that \mathcal{R} is finer than or equal to \mathcal{P} . Then for all $\ell \geq 1$, one has*

$$\begin{aligned} \|\mathbf{E}(f \mid \mathcal{R}) - \mathbf{E}(f \mid \mathcal{P})\|_{S^{2^{\ell}}} &\leq (\mathcal{E}_{f,\ell}(\mathcal{R}_X) - \mathcal{E}_{f,\ell}(\mathcal{P}_X))^{1/2^{\ell}} \\ &\quad + (\mathcal{E}_{f^T,\ell}(\mathcal{R}_Y) - \mathcal{E}_{f^T,\ell}(\mathcal{P}_Y))^{1/2^{\ell}}, \end{aligned}$$

where $f^T : Y \times X \rightarrow \mathbb{C}$ is defined by $f^T(y, x) := f(x, y)$.

Proof. Fix orderings of the elements of X and Y , and let $F := \text{Mat}(f)$. By Lemma 33 and the triangle inequality, the quantity to bound becomes

$$\begin{aligned} \|A_{\mathcal{R}_X} F A_{\mathcal{R}_Y} - A_{\mathcal{P}_X} F A_{\mathcal{P}_Y}\|_{S^{2\ell}} &\leqslant \|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F A_{\mathcal{R}_Y}\|_{S^{2\ell}} + \|A_{\mathcal{P}_X} F (A_{\mathcal{R}_Y} - A_{\mathcal{P}_Y})\|_{S^{2\ell}} \\ &\leqslant \|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F\|_{S^{2\ell}} + \|F (A_{\mathcal{R}_Y} - A_{\mathcal{P}_Y})\|_{S^{2\ell}} \\ &= \|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F\|_{S^{2\ell}} + \|(A_{\mathcal{R}_Y} - A_{\mathcal{P}_Y}) F^T\|_{S^{2\ell}}. \end{aligned}$$

By symmetry, it now suffices to show that

$$\|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F\|_{S^{2\ell}} \leqslant (\mathcal{E}_{F,\ell}(\mathcal{R}_X) - \mathcal{E}_{F,\ell}(\mathcal{P}_X))^{1/2^\ell}.$$

But by (11) and (12), for $1 \leqslant j \leqslant \ell - 1$ we have

$$\begin{aligned} \|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F_j\|_{S^{2\ell+1-j}} &= \|(A_{\mathcal{R}_X} - A_{\mathcal{P}_Y}) F_j F_j^* (A_{\mathcal{R}_X} - A_{\mathcal{P}_X})\|_{S^{2\ell-j}}^{1/2} \\ &\leqslant \|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F_{j+1}\|_{S^{2\ell-j}}^{1/2}. \end{aligned}$$

By inducting on j , this implies that

$$\|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F\|_{S^{2\ell}} \leqslant \|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F_\ell\|_{S^2}^{1/2^{\ell-1}},$$

and we can further write, by (10),

$$\|(A_{\mathcal{R}_X} - A_{\mathcal{P}_X}) F_\ell\|_{S^2}^2 = \|A_{\mathcal{R}_X} F_\ell\|_{S^2}^2 - \|A_{\mathcal{P}_X} F_\ell\|_{S^2}^2 = \mathcal{E}_{F,\ell}(\mathcal{R}) - \mathcal{E}_{F,\ell}(\mathcal{P}).$$

Putting the last two equations together, we obtain our conclusion. \square

3 The regularity lemma

We start by presenting an abstract energy increment argument, the kind of which is typically employed in the proofs of regularity lemmas. Following Tao [49], we will in fact phrase this as an energy optimization argument, which depends on both an *energy function* \mathcal{E} and a *complexity function* \mathcal{C} . This discussion pertains to weak regularity lemmas; a strong regularity lemma would require an additional level of iteration.

Lemma 44 (Abstract energy optimization argument). *Let X be a finite set and \mathcal{P} be a set of partitions of X , which contains the discrete partition $\mathcal{D} := \{\{x\} : x \in X\}$, the trivial partition $\mathcal{T} := \{X\}$, and which is closed under taking common refinements (i.e., if $\mathcal{P}, \mathcal{Q} \in \mathcal{P}$, then $\mathcal{P} \wedge \mathcal{Q} \in \mathcal{P}$). Consider two functions $\mathcal{E}, \mathcal{C} : \mathcal{P} \rightarrow \mathbb{R}$ with the following properties:*

- (i) (Energy is nondecreasing). *For any $\mathcal{P}, \mathcal{P}' \in \mathcal{P}$ such that \mathcal{P}' is finer than \mathcal{P} , one has $\mathcal{E}(\mathcal{P}) \leqslant \mathcal{E}(\mathcal{P}')$.*
- (ii) (Complexity is subadditive). *For any $\mathcal{P}, \mathcal{Q} \in \mathcal{P}$, one has $\mathcal{C}(\mathcal{P} \wedge \mathcal{Q}) \leqslant \mathcal{C}(\mathcal{P}) + \mathcal{C}(\mathcal{Q})$. Moreover, $\mathcal{C}(\mathcal{T}) = 0$.*

Let $\varepsilon > 0$. Then there exists a partition $\mathcal{P} \in \mathcal{P}$ with $\mathcal{C}(\mathcal{P}) \leq \frac{1}{\varepsilon}$, such that for any $\mathcal{Q} \in \mathcal{P}$, one has

$$\mathcal{E}(\mathcal{P} \wedge \mathcal{Q}) - \mathcal{E}(\mathcal{P}) \leq \varepsilon \mathcal{C}(\mathcal{Q}) (\mathcal{E}(\mathcal{D}) - \mathcal{E}(\mathcal{T})).$$

Proof. If $\mathcal{E}(\mathcal{D}) = \mathcal{E}(\mathcal{T})$ (i.e., the energy function is constant), then we can easily pick $\mathcal{P} = \mathcal{T}$, for example. Otherwise, let $\mathcal{P} \in \mathcal{P}$ be a partition which maximizes the real quantity

$$\mathcal{E}(\mathcal{P}) - \varepsilon \mathcal{C}(\mathcal{P}) (\mathcal{E}(\mathcal{D}) - \mathcal{E}(\mathcal{T})).$$

Then by comparing the optimizer \mathcal{P} to the trivial partition \mathcal{T} , we find that

$$\mathcal{E}(\mathcal{T}) \leq \mathcal{E}(\mathcal{P}) - \varepsilon \mathcal{C}(\mathcal{P}) (\mathcal{E}(\mathcal{D}) - \mathcal{E}(\mathcal{T})) \leq \mathcal{E}(\mathcal{D}) - \varepsilon \mathcal{C}(\mathcal{P}) (\mathcal{E}(\mathcal{D}) - \mathcal{E}(\mathcal{T})),$$

and thus $\mathcal{C}(\mathcal{P}) \leq 1/\varepsilon$. Moreover, given any $\mathcal{Q} \in \mathcal{P}$, by comparing \mathcal{P} to $\mathcal{P} \wedge \mathcal{Q}$ we find that

$$\mathcal{E}(\mathcal{P}) - \varepsilon \mathcal{C}(\mathcal{P}) (\mathcal{E}(\mathcal{D}) - \mathcal{E}(\mathcal{T})) \geq \mathcal{E}(\mathcal{P} \wedge \mathcal{Q}) - \varepsilon \mathcal{C}(\mathcal{P} \wedge \mathcal{Q}) (\mathcal{E}(\mathcal{D}) - \mathcal{E}(\mathcal{T})).$$

By the subadditivity of \mathcal{C} , this gives $\mathcal{E}(\mathcal{P} \wedge \mathcal{Q}) - \mathcal{E}(\mathcal{P}) \leq \varepsilon \mathcal{C}(\mathcal{Q}) (\mathcal{E}(\mathcal{D}) - \mathcal{E}(\mathcal{T}))$. \square

Remark 45. Intuitively, Lemma 44 finds a partition \mathcal{P} with a good balance of energy and complexity: it captures most of the energy needed for applications (in the sense that small refinements of \mathcal{P} only see a small increase in energy), while having bounded complexity. Here are some concrete objects that can fulfill the abstract roles of \mathcal{C} , \mathcal{E} and \mathcal{P} from Lemma 44:

- One can take $\mathcal{C}(\mathcal{P})$ to be $\log |\mathcal{P}|$, or the *entropy* $\mathbf{H}(\mathcal{P})$ (by (8)); Tao uses both in [49].
- If Ω is a probability space and $f : \Omega \rightarrow \mathbb{C}$ is an L^2 random variable, one can take $\mathcal{E}(\mathcal{P})$ to be the L^2 energy $\mathbf{E}[\mathbf{E}(f \mid \mathcal{P}_1 \otimes \mathcal{P}_2)^2]$. In this case, one can regard the optimizer $\mathbf{E}(f \mid \mathcal{P}_1 \otimes \mathcal{P}_2)$ as the best approximation to f (of a certain type) with not too many different values.
- Alternatively, given a partition \mathcal{P}_0 of Ω (e.g., induced by a discrete random variable f), one can take $\mathcal{E}(\mathcal{P})$ to be the *mutual information* $\mathbf{I}(\mathcal{P}_0 : \mathcal{P}) = \mathbf{H}(\mathcal{P}_0) + \mathbf{H}(\mathcal{P}) - \mathbf{H}(\mathcal{P}_0 \wedge \mathcal{P})$; this is essentially what Tao does in [49]. In this case, one can regard the optimizer \mathcal{P} as the best approximation to \mathcal{P}_0 (of a certain type) with bounded complexity.
- For a logarithmic and *relative* version of the energy $\mathcal{E}(\mathcal{P})$, see Appendix C.
- If $\Omega = X \times Y$, one can let \mathcal{P} consist of all partitions of the form $\mathcal{P} \otimes \mathcal{Q}$, where \mathcal{P} is a partition of X and \mathcal{Q} is a partition of Y . This is often the case in the proof of regularity lemmas (see Appendix C), but we will actually end up using $\Omega = X$ and $\Omega = Y$ separately.

In our argument, we will use the full space \mathcal{P} of partitions of a probability space (X, \mathbf{P}_X) , both the complexity functions $\log |\mathcal{P}|$ and $\mathbf{H}(\mathcal{P})$, and the generalized L^2 energy functions from Definition 41 (which are monotonic by (10)). To be more precise, we will use a sum of finitely many such energy functions, which resembles a *simultaneous* energy increment argument [51].

Whether it is helpful to use $\log |\mathcal{P}|$ or $\mathbf{H}(\mathcal{P})$ as the complexity function depends on the uniformity norm one wishes to bound. Indeed, while the cut norm of a function $f : X \times Y \rightarrow \mathbb{C}$ depends on the interaction between f and subsets of X and Y (which correspond to partitions with at most two parts), the *spectral norm* S^∞ can be characterized in terms of the interaction between f and partitions with bounded entropy; this is the content of the following lemma.

Lemma 46 (Spectral norms and entropy). *Given a finite probability space (Z, \mathbf{P}_Z) , denote*

$$\mathcal{F}(Z) := \{g : Z \rightarrow \{0\} \cup \{2^k, i2^k, -2^k, -i2^k : k \in \mathbb{Z}_{\geq 0}\} \mid \|g\|_{L^2} \leq 3\}. \quad (13)$$

Then any $g \in \mathcal{F}(Z)$ has $\mathbf{H}(g) \ll 1$. Moreover, for any finite probability spaces (X, \mathbf{P}_X) , (Y, \mathbf{P}_Y) , and $f : X \times Y \rightarrow \mathbb{C}$, one has

$$\|f\|_{S^\infty} \leq \sup_{\substack{g \in \mathcal{F}(X) \\ h \in \mathcal{F}(Y)}} |\mathbf{E}_{x,y}[f(x, y)g(x)h(y)]|.$$

Proof. We leave this computation to Appendix A. □

We are now ready to state and prove our sparse weak regularity lemma, which generalizes Theorem 4. For technical reasons, we state our result for a finite family of $k \geq 1$ functions rather than a single function $f : X \times Y \rightarrow \mathbb{C}$.

Theorem 47 (Simultaneous weak regularity with Schatten norms). *Let $\{(X_i, \mathbf{P}_i)\}_{1 \leq i \leq m}$ be finite probability spaces, $1 \leq q_t < r_t \leq \infty$, $i_t, j_t \in \{1, \dots, m\}$ and $f_t : X_{i_t} \times X_{j_t} \rightarrow \mathbb{C}$ for $t \in \{1, \dots, k\}$, and $\varepsilon \in (0, 1)$. Then there exist partitions $\mathcal{P}_i, \mathcal{P}'_i$ of X_i such that*

$$\max_i \log |\mathcal{P}_i| \ll_{k,q} \frac{1}{\varepsilon^{\alpha(q,\infty)}}, \quad \max_t \|f_t - \mathbf{E}(f_t \mid \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t})\|_{\square} \leq \varepsilon \max_t \|f_t\|_{S^{q_t}}, \quad (14)$$

respectively

$$\max_i \mathbf{H}(\mathcal{P}'_i) \ll_{k,q,r} \frac{1}{\varepsilon^{\alpha(q,r)}}, \quad \max_t \|f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^{r_t}} \leq \varepsilon \max_t \|f_t\|_{S^{q_t}}, \quad (15)$$

where $q := \max_t q_t$, $r := q \min_t (r_t/q_t)$, and $\alpha(q, r) := 2^{\max(\lceil \log_2 q \rceil, 1)} / (1 - \frac{q}{r})$.

Remark 48. Ultimately, we will only use Theorem 47 for functions f with values in $[0, \infty)$, but it is not much harder to treat all complex-valued functions. Also, our sparse regularity method will only use (15), but the cut norm bound in (14) may be of independent interest to the reader (e.g., since it fits better with the counting lemmas of [14, 15]).

While Schatten norms are stronger than cut norms, the drawback of using (15) is that it only controls the entropies $\mathbf{H}(\mathcal{P}'_i)$ (which could be much smaller than $\log |\mathcal{P}'_i|$). However, this disadvantage can be removed if one is allowed to disregard a small number of elements of each X_i , which is the case for most applications.

Proof of Theorem 47. We first note that it suffices to consider the case when all $r_t = \infty$, since Schatten norms can be interpolated. Indeed, assuming that (15) holds when all $r_t = \infty$, one can apply it for $2^{1/(1-\frac{r}{q})}\varepsilon^{1/(1-\frac{q}{r})}$ in place of ε (for any given $r \in (q, \infty)$); this yields partitions \mathcal{P}'_i with

$$\max_i \mathbf{H}(\mathcal{P}'_i) \ll_q \frac{1}{2^{\alpha(q,\infty)/(1-\frac{r}{q})} \varepsilon^{\alpha(q,\infty)/(1-\frac{q}{r})}} \ll_{q,r} \frac{1}{\varepsilon^{\alpha(q,r)}},$$

such that

$$\max_t \|f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^\infty} \leq 2^{1/(1-\frac{r}{q})} \varepsilon^{1/(1-\frac{q}{r})} \max_t \|f_t\|_{S^{q_t}}.$$

But then, given any values $r_t \in (q_t, \infty]$ and setting $r := q \min_t (r_t/q_t)$, Lemma 37 implies that if $M := \max_t \|f_t\|_{S^{q_t}}$, then for each $t \in \{1, \dots, k\}$ one has

$$\begin{aligned} & \|f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^{r_t}} \\ & \leq \|f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^{q_t}}^{q_t/r_t} \cdot \|f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^\infty}^{1-q_t/r_t} \\ & \leq (\|f_t\|_{S^{q_t}} + \|\mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^{q_t}})^{q_t/r_t} \cdot \left(2^{1/(1-\frac{r}{q})} \varepsilon^{1/(1-\frac{q}{r})} M\right)^{1-q_t/r_t} \\ & \leq (2M)^{q_t/r_t} \cdot 2^{-q_t/r_t} \varepsilon M^{1-q_t/r_t} \\ & = \varepsilon M. \end{aligned}$$

After this reduction, it remains to find partitions \mathcal{P}_i which satisfy (14), and \mathcal{P}'_i which satisfy (15) for $r = r_t = \infty$. For now, fix some partitions \mathcal{P}_i and \mathcal{P}'_i of X_i , which will be chosen later via the abstract energy optimization argument in Lemma 44; it will take some work to explain how exactly this optimization argument will be applied. By (1) and Lemma 46, we can express the uniformity norms of interest (for $t \in \{1, \dots, k\}$) as

$$\begin{aligned} & \|f_t - \mathbf{E}(f_t \mid \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t})\|_\square = \sup_{\substack{g_1: X_{i_t} \rightarrow \{0,1\} \\ g_2: X_{j_t} \rightarrow \{0,1\}}} |\mathbf{E}_{x_{1,2}}[(f_t - \mathbf{E}(f_t \mid \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t}))(x_1, x_2)g_1(x_1)g_2(x_2)]|, \\ & \|f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^\infty} \leq \sup_{\substack{h_1 \in \mathcal{F}(X_{i_t}) \\ h_2 \in \mathcal{F}(X_{j_t})}} |\mathbf{E}_{x_{1,2}}[(f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t}))(x_1, x_2)h_1(x_1)h_2(x_2)]|, \end{aligned}$$

where x_1 is sampled from $(X_{i_t}, \mathbf{P}_{i_t})$, x_2 is sampled from $(X_{j_t}, \mathbf{P}_{j_t})$, and $\mathcal{F}(X_i)$ are as in (13). In particular, all functions involved in both suprema have $\|g_u\|_{L^2} \ll 1$ and $\|h_u\|_{L^2} \ll 1$, while each g_u in the first supremum takes at most two values, and each h_u in the second supremum has $\mathbf{H}(h_u) \ll 1$ (for $u \in \{1, 2\}$). Taking a maximum over t and suppressing dependencies on x_1, x_2 for notational simplicity, we get

$$\begin{aligned} & \max_{t \in \{1, \dots, k\}} \|f_t - \mathbf{E}(f_t \mid \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t})\|_\square = \sup_{\substack{g_1^t: X_{i_t} \rightarrow \{0,1\} \\ g_2^t: X_{j_t} \rightarrow \{0,1\} \\ t \in \{1, \dots, k\}}} |\mathbf{E}[(f_t - \mathbf{E}(f_t \mid \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t}))g_1^t g_2^t]|, \\ & \max_{t \in \{1, \dots, k\}} \|f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^\infty} \leq \sup_{\substack{h_1^t \in \mathcal{F}(X_{i_t}) \\ h_2^t \in \mathcal{F}(X_{j_t}) \\ t \in \{1, \dots, k\}}} |\mathbf{E}[(f_t - \mathbf{E}(f_t \mid \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t}))h_1^t h_2^t]|, \end{aligned}$$

Thus it suffices to bound the right-hand sides for any specific choice of the tuples of functions $\{g_u^t\}_{t,u}$, respectively $\{h_u^t\}_{t,u}$. Fix such tuples of functions, and consider the partitions of X_i that they collectively induce:

$$\mathcal{Q}_i := \bigvee_{\substack{t \in \{1, \dots, k\} \\ i_t = i}} \text{Part}(g_1^t) \wedge \bigvee_{\substack{t \in \{1, \dots, k\} \\ j_t = i}} \text{Part}(g_2^t), \quad \mathcal{Q}'_i := \bigvee_{\substack{t \in \{1, \dots, k\} \\ i_t = i}} \text{Part}(h_1^t) \wedge \bigvee_{\substack{t \in \{1, \dots, k\} \\ j_t = i}} \text{Part}(h_2^t).$$

Due to the nature of the functions g_u^t and h_u^t , these partitions satisfy the the complexity bounds

$$\log |\mathcal{Q}_i| \ll_k 1 \quad \text{and} \quad \mathbf{H}(\mathcal{Q}'_i) \ll_k 1. \quad (16)$$

For notational convenience, we also denote

$$\mathcal{R}_i := \mathcal{P}_i \wedge \mathcal{Q}_i \quad \text{and} \quad \mathcal{R}'_i := \mathcal{P}'_i \wedge \mathcal{Q}'_i, \quad (17)$$

which begins to resemble the setting of our energy optimization argument (Lemma 44). Now for $t \in \{1, \dots, k\}$, using (4) and the fact that the function $(x_1, x_2) \mapsto g_1^{i,j}(x_1)g_2^{i,j}(x_2)$ is $(\mathcal{R}_{i_t} \otimes \mathcal{R}_{j_t})$ -measurable, we can write

$$\begin{aligned} & |\mathbf{E}[(f_t - \mathbf{E}(f_t | \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t}))g_1^t g_2^t]| \\ &= |\mathbf{E}[(\mathbf{E}(f_t | \mathcal{R}_{i_t} \otimes \mathcal{R}_{j_t}) - \mathbf{E}(f_t | \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t}))g_1^t g_2^t]| \\ &\leq \|\mathbf{E}(f_t | \mathcal{R}_{i_t} \otimes \mathcal{R}_{j_t}) - \mathbf{E}(f_t | \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t})\|_{S^\infty} \cdot \|g_1^t\|_{L^2} \cdot \|g_2^t\|_{L^2} \\ &\ll \|\mathbf{E}(f_t | \mathcal{R}_{i_t} \otimes \mathcal{R}_{j_t}) - \mathbf{E}(f_t | \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t})\|_{S^\infty}, \end{aligned}$$

and similarly,

$$|\mathbf{E}[(f_t - \mathbf{E}(f_t | \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t}))h_1^t h_2^t]| \ll \|\mathbf{E}(f_t | \mathcal{R}'_{i_t} \otimes \mathcal{R}'_{j_t}) - \mathbf{E}(f_t | \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^\infty}.$$

Putting things together, to obtain the desired uniformity norm bounds in (14) and (15), it suffices to have

$$\max_t \|\mathbf{E}(f_t | \mathcal{R}_{i_t} \otimes \mathcal{R}_{j_t}) - \mathbf{E}(f_t | \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t})\|_{S^\infty} \leq \varepsilon \max_t \|f_t\|_{S^q}, \quad (18)$$

$$\max_t \|\mathbf{E}(f_t | \mathcal{R}'_{i_t} \otimes \mathcal{R}'_{j_t}) - \mathbf{E}(f_t | \mathcal{P}'_{i_t} \otimes \mathcal{P}'_{j_t})\|_{S^\infty} \leq \varepsilon \max_t \|f_t\|_{S^q}. \quad (19)$$

We will first present the argument for (18), and then mention what changes for (19). By Lemma 43 for $\ell := \max(\lceil \log_2 q \rceil, 1) \geq \log_2 q$, for each t we have

$$\begin{aligned} \|\mathbf{E}(f_t | \mathcal{R}_{i_t} \otimes \mathcal{R}_{j_t}) - \mathbf{E}(f_t | \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t})\|_{S^\infty} &\leq (\mathcal{E}_{f_t, \ell}(\mathcal{R}_{i_t}) - \mathcal{E}_{f_t, \ell}(\mathcal{P}_{i_t}))^{1/2^\ell} \\ &\quad + (\mathcal{E}_{f_t^T, \ell}(\mathcal{R}_{j_t}) - \mathcal{E}_{f_t^T, \ell}(\mathcal{P}_{j_t}))^{1/2^\ell}, \end{aligned} \quad (20)$$

where all energies are bounded by $\mathcal{E}_{f_t, \ell} = \|f_t\|_{S^{2^\ell}}^{2^\ell} \leq \|f_t\|_{S^q}^{2^\ell} \leq \|f_t\|_{S^{q_t}}^{2^\ell}$. We are finally ready to choose \mathcal{P}_i by applying the energy optimization argument in Lemma 44, once for each $i \in \{1, \dots, m\}$, to the set $\Omega = X_i$, in the search space \mathcal{P} of all partitions of X_i , with the

complexity function $\log |\cdot|$, $c\varepsilon^{2^\ell}$ in place of ε (for some $c > 0$ to be chosen shortly), and with the energy function

$$\mathcal{E}_i(\mathcal{P}_i) := \sum_{\substack{t \in \{1, \dots, k\} \\ i_t = i}} \mathcal{E}_{f_t, \ell}(\mathcal{P}_i) + \sum_{\substack{t \in \{1, \dots, k\} \\ j_t = i}} \mathcal{E}_{f_t^T, \ell}(\mathcal{P}_i).$$

Recalling the notation $M := \max_t \|f_t\|_{S^{q_t}}$, note that the energy function \mathcal{E}_i is nonnegative and bounded above by

$$\sum_{\substack{t \in \{1, \dots, k\} \\ i_t = i}} \mathcal{E}_{f_t, \ell} + \sum_{\substack{t \in \{1, \dots, k\} \\ j_t = i}} \mathcal{E}_{f_t^T, \ell} \leq \sum_{\substack{t \in \{1, \dots, k\} \\ i_t = i}} \|f_t\|_{S^{q_t}}^{2^\ell} + \sum_{\substack{t \in \{1, \dots, k\} \\ j_t = i}} \|f_t\|_{S^{q_t}}^{2^\ell} \leq 2kM^{2^\ell}.$$

Thus the resulting \mathcal{P}_i satisfy

$$\log |\mathcal{P}_i| \ll \frac{1}{c\varepsilon^{2^\ell}} \ll_c \frac{1}{\varepsilon^{\alpha(q, \infty)}}$$

and, due to the conclusion of Lemma 44, (17) and (16),

$$\mathcal{E}_i(\mathcal{R}_i) - \mathcal{E}_i(\mathcal{P}_i) \ll_k c\varepsilon^{2^\ell} \cdot 2kM^{2^\ell} \cdot \log |\mathcal{Q}_i| \ll_k c(\varepsilon M)^{2^\ell}.$$

Choosing $c = c(k, q) > 0$ small enough in terms of k and $\ell = \ell(q)$, we conclude that

$$\sum_{\substack{t \in \{1, \dots, k\} \\ i_t = i}} (\mathcal{E}_{f_t, \ell}(\mathcal{R}_i) - \mathcal{E}_{f_t, \ell}(\mathcal{P}_i)) + \sum_{\substack{t \in \{1, \dots, k\} \\ j_t = i}} (\mathcal{E}_{f_t^T, \ell}(\mathcal{R}_i) - \mathcal{E}_{f_t^T, \ell}(\mathcal{P}_i)) \leq \left(\frac{\varepsilon M}{2}\right)^{2^\ell},$$

where each term in the two sums is nonnegative. Putting this together with (20), we obtain

$$\|\mathbf{E}(f_t \mid \mathcal{R}_{i_t} \otimes \mathcal{R}_{j_t}) - \mathbf{E}(f_t \mid \mathcal{P}_{i_t} \otimes \mathcal{P}_{j_t})\|_{S^\infty} \leq 2^{-1}\varepsilon M + 2^{-1}\varepsilon M = \varepsilon M,$$

and taking a maximum over $t \in \{1, \dots, k\}$ yields the desired bound in (18).

To establish (19), we choose the partitions \mathcal{P}'_i via Lemma 44 in the same way, except that we use the complexity function $\mathbf{H}(\cdot)$ instead of $\log |\cdot|$. This yields

$$\mathbf{H}(\mathcal{P}'_i) \ll_c \frac{1}{\varepsilon^{\alpha(q, \infty)}} \quad \text{and} \quad \mathcal{E}_i(\mathcal{R}'_i) - \mathcal{E}_i(\mathcal{P}'_i) \ll_k c\varepsilon^{2^\ell} \cdot 2kM^{2^\ell} \cdot \mathbf{H}(\mathcal{Q}'_i),$$

and choosing a suitable constant $c = c(k, q)$ is now possible due to the entropy bound in (16). \square

Remark 49. One can also give a constructive proof of a statement like Theorem 47 using singular value decomposition (related to the spectral methods of Frieze and Kannan [25, 24]), at the expense of inserting cutoffs of the form $\mathbb{1}_{R_i^c \times R_j^c}$ in the uniformity norms, for some small exceptional sets R_i . Such cutoffs are compatible with our methods, and can eliminate the need to use entropy as a measure of complexity. We chose the method of proof above partly since we find it cleaner, and partly since the generalized L^2 energies from Definition 41 might find other applications (in particular, they may have higher-order combinatorial analogues, applicable in the setting of hypergraphs).

4 The counting lemma

For comparison purposes, we first state a version of the k -cycle counting lemma for dense weighted graphs (see a similar version for triangles, e.g., in [14, Lemma 3.4]).

Lemma 50 (Dense cycle counting lemma). *Let $k \in \mathbb{Z}_{\geq 3}$ and $\varepsilon, C > 0$. Let (X_i, \mathbf{P}_i) be finite probability spaces and $f_i, \tilde{f}_i : X_i \times X_{i+1} \rightarrow [0, \infty)$ be functions satisfying $\|f_i\|_{L^\infty} \leq C$, $\|\tilde{f}_i\|_{L^\infty} \leq C$, and $\|f_i - \tilde{f}_i\|_{\square} \leq \varepsilon$, for $i \in \mathbb{Z}/k\mathbb{Z}$. Then*

$$\left| \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) - \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \tilde{f}_i(x_i, x_{i+1}) \right| \leq kC^{k-1}\varepsilon.$$

Like in the case of regularity lemmas, a *sparse* counting lemma should not require L^∞ bounds on the functions f_i . The counting lemmas of [17, 14, 15], for instance, assume that \tilde{f}_i remain L^∞ -bounded, but have weaker assumptions about f_i . In our sparse counting lemma below, we eliminate all of the L^∞ constraints, by replacing them with appropriate Schatten norm bounds. This weakening of the *density norm* bounds comes at the expense of stronger *uniformity norm* bounds, as we also replace the cut norm by the appropriate Schatten norms.

Lemma 51 (Sparse cycle counting lemma with Schatten norms). *Let $k \in \mathbb{Z}_{\geq 2}$, $q_1, \dots, q_k \in [1, \infty)$, and $r_1, \dots, r_k \in (1, \infty]$ be such that for each $i \in \mathbb{Z}/k\mathbb{Z}$,*

$$\frac{1}{r_i} + \sum_{j \neq i} \frac{1}{q_j} \geq 1.$$

Also, let $\varepsilon, C > 0$, (X_i, \mathbf{P}_i) be finite probability spaces, and $f_i, \tilde{f}_i : X_i \times X_{i+1} \rightarrow \mathbb{C}$ be functions satisfying $\|f_i\|_{S^{q_i}} \leq C$, $\|\tilde{f}_i\|_{S^{q_i}} \leq C$, and $\|f_i - \tilde{f}_i\|_{S^{r_i}} \leq \varepsilon$, for $i \in \mathbb{Z}/k\mathbb{Z}$. Then

$$\left| \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) - \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \tilde{f}_i(x_i, x_{i+1}) \right| \leq kC^{k-1}\varepsilon.$$

Proof. Replacing f_i with \tilde{f}_i one i at a time, it is enough to show (by the circular symmetry of the indices and the equivalent conditions on f_i and \tilde{f}_i) that

$$\left| \mathbf{E}_{x_1, \dots, x_k} (f_1 - \tilde{f}_1)(x_1, x_2) \prod_{i=2}^k f_i(x_i, x_{i+1}) \right| \leq C^{k-1}\varepsilon.$$

But is an immediate consequence of the cycle counting inequality from Lemma 37, our assumed Schatten norm bounds, and the fact that $\frac{1}{r_i} + \sum_{j \neq i} \frac{1}{q_j} \geq 1$. \square

Remark 52. The counting lemmas considered in [14, 15] are one-sided, in the sense that they only provide lower bounds for expressions like $\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1})$. Conlon, Fox, Sudakov and Zhao also asked [15] about potential two-sided results: specifically,

for which (connected) subgraphs F is there a two-sided F -counting lemma, in graphs G with no 4-cycles? Lemma 51 answers this question when F is a cycle or a path, in the broader setting of graphs G with not too many 4-cycles; the only drawback is that it uses the slightly stronger S^∞ uniformity norm instead of the cut norm. The reader should compare the corollary below with the notion of *countable* graphs from [15].

Corollary 53 (Two-sided counting lemma in graphs with not too many C_4 's). *For any graph $F \in \{C_k : k \geq 5\} \cup \{P_k : k \geq 1\}$ and any $\varepsilon, C > 0$, there exists $\delta > 0$ such that the following holds. Let G be an n -vertex graph with $\text{Inj}(C_4, G) \leq Cn^2$, and $H \in [0, \infty)^{V(G) \times V(G)}$ be a symmetric matrix with $\text{Tr}(H^4) \leq Cn^4$, such that*

$$\forall a, b : V(G) \rightarrow \mathbb{C} \quad \text{with} \quad \|a\|_2, \|b\|_2 \leq \sqrt{n},$$

$$\left| \frac{\sum_{x,y} \mathbb{1}_G(x, y) a(x) b(y)}{n^{3/2}} - \frac{\sum_{x,y} H(x, y) a(x) b(y)}{n^2} \right| \leq \delta.$$

Then for every family $(A_v)_{v \in V(F)}$ of subsets $A_v \subset V(G)$, one has

$$\left| \frac{\sum_{x_v \in A_v, \forall v \in V(F)} \prod_{\{u,v\} \in E(F)} \mathbb{1}_G(x_u, x_v)}{n^{|V(F)|-|E(F)|/2}} - \frac{\sum_{x_v \in A_v, \forall v \in V(F)} \prod_{\{u,v\} \in E(F)} H(x_u, x_v)}{n^{|V(F)|}} \right| \leq \varepsilon.$$

Proof. Equip $V(G)$ with the uniform probability measure, and define $f, \tilde{f} : V(G) \times V(G) \rightarrow [0, \infty)$ by $f := \mathbb{1}_G/p$ and $\tilde{f} := H$, for $p = n^{-1/2}$ (one can in fact use any $p \geq n^{-1/2}$ to obtain a slightly more general result). A short computation of Conlon, Fox, Sudakov and Zhao [14, p. 2389, below (11)] shows that for any n -vertex graph G , one has

$$\text{Hom}(C_4, G) \ll n^2 + \text{Inj}(C_4, G), \quad (21)$$

with an absolute implied constant. Using (5) and our assumption that $\text{Inj}(C_4, G) \leq Cn^4$, this yields

$$\|f\|_{S^4}^4 = \frac{\text{Hom}(C_4, G)}{p^4 n^4} \leq \frac{C' n^2}{p^4 n^4} = C',$$

for some constant $C' > 0$ (depending only on C). Similarly, $\text{Tr}(H^4) \leq Cn^4$ translates to $\|\tilde{f}\|_{S^4} \leq C^{1/4}$, and the assumed δ -bound involving functions $a, b : V(G) \rightarrow \mathbb{C}$ simply states that

$$\|f - \tilde{f}\|_{S^\infty} \leq \delta,$$

by (4) (in [15], a and b are required to be indicator functions of subsets of $V(G)$, which amounts to the less restrictive cut norm bound $\|f - \tilde{f}\|_\square \leq \delta$). Finally, the desired conclusion translates to

$$\left| \mathbf{E}_{x_1, \dots, x_k} \prod_{v \in \mathbb{Z}/k\mathbb{Z}} (f \mathbb{1}_{A_v \times A_{v+1}})(x_v, x_{v+1}) - \mathbf{E}_{x_1, \dots, x_k} \prod_{v \in \mathbb{Z}/k\mathbb{Z}} (\tilde{f} \mathbb{1}_{A_v \times A_{v+1}})(x_v, x_{v+1}) \right| \leq \varepsilon, \quad (22)$$

when $F = C_k$ is a k -cycle with $k \geq 5$ (and $V(F) = \mathbb{Z}/k\mathbb{Z}$), respectively

$$\left| \mathbf{E}_{x_1, \dots, x_k} \prod_{v=1}^{k-1} (f \mathbb{1}_{A_v \times A_{v+1}})(x_v, x_{v+1}) - \mathbf{E}_{x_1, \dots, x_k} \prod_{v=1}^{k-1} (\tilde{f} \mathbb{1}_{A_v \times A_{v+1}})(x_v, x_{v+1}) \right| \leq \varepsilon, \quad (23)$$

when $F = P_{k-1}$ is a $(k-1)$ -path with $k \geq 2$ (and vertex set $V(F) = \{1, \dots, k\}$). Letting $f_v := f \mathbb{1}_{A_v \times A_{v+1}}$ and $\tilde{f}_v := \tilde{f} \mathbb{1}_{A_v \times A_{v+1}}$, we still have $\|f_v\|_{S^4}, \|\tilde{f}_v\|_{S^4} \ll_C 1$ and $\|f_v - \tilde{f}_v\|_{S^\infty} \leq \delta$ for each v , by Lemma 37.

Thus (22) follows immediately from Lemma 51, by picking a small enough δ and noting that $\frac{1}{\infty} + \sum_{v=1}^{k-1} \frac{1}{4} \geq 1$. Similarly, (23) follows from Lemma 51, by inserting constant functions $f_k = \tilde{f}_k := \mathbb{1}_{V(G)}$ with $\|f_k\|_{S^1} = \|\tilde{f}_k\|_{S^1} = 1$ (by Lemma 37), and noting that $\frac{1}{\infty} + \sum_{v=1}^{k-2} \frac{1}{4} + 1 \geq 1$. \square

Our next lemma is a bit technical, but it will ultimately allow us to consider only proper k -cycles (as opposed to homomorphic copies of k -cycles) in results about non- k -partite graphs such as Corollary 15. This is done by following an idea from [14] to carefully partition the space X into k parts $X^{(1)}, \dots, X^{(k)}$, and to only consider the tuples $(x_1, \dots, x_k) \in X^{(1)} \times \dots \times X^{(k)}$.

Lemma 54 (Restricting to distinct vertices). *Let $k \in \mathbb{Z}_{\geq 2}$, (X, \mathbf{P}) be a finite space equipped with uniform probability, and $f : X^k \rightarrow [0, \infty)$ be a $(\bigotimes_{i=1}^k \mathcal{P})$ -measurable function, for some partition \mathcal{P} of X . If R denotes the union of all parts of \mathcal{P} with fewer than $k(k+1)$ elements, then there exists a partition of X into k disjoint parts $X^{(1)}, \dots, X^{(k)}$ such that*

$$\mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) \prod_{i=1}^k \mathbb{1}_{R^c}(x_i) \ll_k \mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) \prod_{i=1}^k \mathbb{1}_{X^{(i)}}(x_i).$$

Proof. We split each part $A \in \mathcal{P}$ arbitrarily into k (possibly empty) parts of sizes differing by at most 1, labeled by $A^{(i)}$ for $i \in \mathbb{Z}/k\mathbb{Z}$, and let $X^{(i)}$ be the union of all such $A^{(i)}$.

Now let A_1, \dots, A_k denote some (not necessarily distinct) parts of \mathcal{P} , which have at least $k(k+1)$ vertices (i.e., $A_1, \dots, A_k \subset R^c$). Consider the expectation

$$\mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) \prod_{i=1}^k \mathbb{1}_{X^{(i)} \cap A_i}(x_i),$$

where x_1, \dots, x_k still range over X , although only the values $(x_1, \dots, x_k) \in A_1 \times \dots \times A_k$ are relevant. Note that for these values, $f(x_1, \dots, x_k)$ is equal to some nonnegative constant $c(A_1, \dots, A_k)$. Moreover, by the construction of $X^{(i)}$, the intersection $X^{(i)} \cap A_i$ contains at least

$$\left\lfloor \frac{|A_i|}{k} \right\rfloor \geq \frac{|A_i| - k}{k} \geq \frac{|A_i|}{k+1}$$

vertices, and thus $\mathbf{E}_{x_i} \mathbb{1}_{X^{(i)} \cap A_i}(x_i) \geq \frac{1}{k+1} \mathbf{E}_{x_i} \mathbb{1}_{A_i}(x_i)$ (since we are using the uniform probability distribution on X). It follows that

$$\begin{aligned} \mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) \prod_{i=1}^k \mathbb{1}_{X^{(i)} \cap A_i}(x_i) &= c(A_1, \dots, A_k) \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbb{1}_{X^{(i)} \cap A_i}(x_i) \\ &\geq \frac{1}{(k+1)^k} c(A_1, \dots, A_k) \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbb{1}_{A_i}(x_i) \\ &= \frac{1}{(k+1)^k} \mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) \prod_{i=1}^k \mathbb{1}_{A_i}(x_i), \end{aligned}$$

and summing over all choices of $A_1 \times \dots \times A_k$, we conclude that

$$\begin{aligned} \mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) \prod_{i=1}^k \mathbb{1}_{R^c}(x_i) &\ll_k \mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) \prod_{i=1}^k \mathbb{1}_{X^{(i)} \cap R^c}(x_i) \\ &\leq \mathbf{E}_{x_1, \dots, x_k} f(x_1, \dots, x_k) \prod_{i=1}^k \mathbb{1}_{X^{(i)}}(x_i), \end{aligned}$$

using the nonnegativity of f once again. \square

We can now present the natural consequence of our regularity and counting lemmas (with no transference-step machinery or assumptions involved), which reduces a general k -cycle count in a weighted k -partite graph to an analogous count where the weight functions are fairly structured. When all vertex sets are equal, Lemma 54 also allows us to consider strictly proper k -cycles (with distinct vertices) in the upper bound.

Proposition 55 (Reduction to low complexity). *For any $k \in \mathbb{Z}_{\geq 2}$ and $q_1, \dots, q_k \in [1, \infty)$ with $\sum_i \frac{1}{q_i} > 1$, and any $\varepsilon, C > 0$, there exists $\tau > 0$ such that the following hold. Let (X_i, \mathbf{P}_i) be finite probability spaces, and $f_i : X_i \times X_{i+1} \rightarrow \mathbb{C}$ satisfy $\|f_i\|_{S^{q_i}} \leq C$, for $i \in \mathbb{Z}/k\mathbb{Z}$.*

(i). *There exist partitions \mathcal{P}'_i of X_i with $\mathbf{H}(\mathcal{P}'_i) = O_{\varepsilon, k, q_1, \dots, q_k, C}(1)$, such that*

$$\left| \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) - \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P}'_i \otimes \mathcal{P}'_{i+1})(x_i, x_{i+1}) \right| \leq \varepsilon.$$

(ii). *If all f_i take values in $[0, \infty)$, there exist partitions $\mathcal{P}_i = \{R_i, A_{i,1}, A_{i,2}, \dots\}$ of X_i , with $\mathbf{P}_i(R_i) \leq \varepsilon^2$ and $\mathbf{P}(A_{i,j}) \geq \tau$ for all i, j , such that*

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P}_i \otimes \mathcal{P}_{i+1})(x_i, x_{i+1}) \mathbb{1}_{R_i^c}(x_i) \leq \varepsilon + \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}).$$

(iii). If additionally all spaces $(X_i, \mathbf{P}_i) = (X, \mathbf{P})$ are equal and \mathbf{P} is the uniform probability distribution, and $|X|$ is sufficiently large in terms of $\varepsilon, k, q_1, \dots, q_k, C$, then in part (ii) one can take all $\mathcal{P}_i = \mathcal{P}$ and $R_i = R$ to be equal, and also conclude the stronger bound

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P} \otimes \mathcal{P})(x_i, x_{i+1}) \mathbb{1}_{R^c}(x_i) \ll_k \varepsilon + \mathbf{E}_{x_1, \dots, x_k} \mathbb{1}_{x_1, \dots, x_k \text{ dist.}} \prod_{i=1}^k f_i(x_i, x_{i+1}).$$

Remark 56. Parts (ii) and (iii) of Proposition 55 yield partitions \mathcal{P}_i whose parts all have probabilities at least τ (except possibly for one part R_i), which automatically implies the cardinality bound $|\mathcal{P}_i| \leq \frac{1}{\tau} + 1 = O_{\varepsilon, k, q_1, \dots, q_k, C}(1)$. This is already better than the entropy bound $\mathbf{H}(\mathcal{P}'_i) = O_{\varepsilon, k, q_1, \dots, q_k, C}(1)$ from part (i), but the more explicit restriction on the structure of \mathcal{P}_i will be useful in Section 6.

Proof. Assume without loss of generality that $\varepsilon/(kC^k) < 1$ (by making ε smaller if necessary). To prove (i), apply Theorem 47 for the functions f_i , and with

$$\begin{aligned} r_i := q_i + q_i^2 \left(\sum_j \frac{1}{q_j} - 1 \right) > q_i \quad \Rightarrow \quad \frac{1}{q_i} - \frac{1}{r_i} &\leq \frac{r_i - q_i}{q_i^2} = \sum_j \frac{1}{q_j} - 1 \\ &\Rightarrow \quad \frac{1}{r_i} + \sum_{j \neq i} \frac{1}{q_j} \geq 1. \end{aligned} \tag{24}$$

This produces partitions \mathcal{P}'_i of X_i such that $\mathbf{H}(\mathcal{P}'_i) = O_{\varepsilon, k, q_1, \dots, q_k, C}(1)$ and

$$\max_i \|f_i - \mathbf{E}(f_i \mid \mathcal{P}'_i \otimes \mathcal{P}'_{i+1})\|_{S^{r_i}} \leq \frac{\varepsilon}{kC^k} \max_i \|f_i\|_{S^{q_i}} \leq \frac{\varepsilon}{kC^{k-1}}.$$

The conclusion now follows from our counting lemma (Lemma 51).

For (ii), consider the partitions \mathcal{P}'_i from (i), and let \mathcal{P}_i be obtained from \mathcal{P}'_i by combining all parts with probability less than τ into one part R_i (for some parameter $\tau \in (0, 1)$ to be chosen shortly). Then, if p_1, p_2, \dots denote the probabilities of all parts of \mathcal{P}'_i , we have

$$\begin{aligned} \mathbf{P}_i(R_i) &= \sum_{p_n < \tau} p_n = \frac{1}{\log(1/\tau)} \sum_{p_n < \tau} p_n \log \frac{1}{\tau} \\ &\leq \frac{1}{\log(1/\tau)} \sum_n p_n \log \frac{1}{p_n} = \frac{\mathbf{H}(\mathcal{P}'_i)}{\log(1/\tau)} = O_{\varepsilon, k, q, C} \left(\frac{1}{\log(1/\tau)} \right). \end{aligned} \tag{25}$$

We can thus pick $\tau = \tau(\varepsilon, k, q_1, \dots, q_k, C)$ small enough so that $\mathbf{P}_i(R_i) \leq \varepsilon^2$ (uniformly in i). Since the functions $\mathbf{E}(f_i \mid \mathcal{P}'_i \otimes \mathcal{P}'_{i+1})$ and $\mathbf{E}(f_i \mid \mathcal{P}_i \otimes \mathcal{P}_{i+1})$ are equal inside $R_i^c \times R_{i+1}^c$, we conclude (using (i)) that

$$\begin{aligned} \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P}_i \otimes \mathcal{P}_{i+1}) \mathbb{1}_{R_i^c}(x_i) &\leq \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P}'_i \otimes \mathcal{P}'_{i+1})(x_i, x_{i+1}) \\ &\leq \varepsilon + \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}). \end{aligned}$$

Finally, to prove (iii), we apply Theorem 47 as before, except that now all vertex sets are equal (so $m = 1$). This produces a partition \mathcal{P}' of X such that $\mathbf{H}(\mathcal{P}') = O_{\varepsilon, k, q_1, \dots, q_k, C}(1)$ and

$$\max_i \|f_i - \mathbf{E}(f_i \mid \mathcal{P}' \otimes \mathcal{P}')\|_{S^{r_i}} \leq \frac{\varepsilon}{kC^k} \max_i \|f_i\|_{S^{q_i}} \leq \frac{\varepsilon}{kC^{k-1}}.$$

We repeat the construction of \mathcal{P}_i and R_i from part (ii) (except that now there is only one partition), to obtain a partition \mathcal{P} with $|\mathcal{P}| = O_{\varepsilon, k, q_1, \dots, q_k, C}(1)$, and a \mathcal{P} -measurable set R with $\mathbf{P}(R) \leq \varepsilon^2$, which is the union of all parts of \mathcal{P}' with probability less than $\tau = \tau(\varepsilon, k, q_1, \dots, q_k, C)$.

Now assume that $|X|$ is large enough so that $k(k+1)/|X| < \tau$, and apply Lemma 54 for the partition \mathcal{P}' and the $(\bigotimes_{i=1}^k \mathcal{P}')$ -measurable function

$$f(x_1, \dots, x_k) := \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P}' \otimes \mathcal{P}')(x_i, x_{i+1}),$$

to obtain a partition of X into k disjoint parts $X^{(1)}, \dots, X^{(k)}$, such that

$$\begin{aligned} & \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P}' \otimes \mathcal{P}')(x_i, x_{i+1}) \mathbb{1}_{R^c}(x_i) \\ & \ll_k \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P}' \otimes \mathcal{P}')(x_i, x_{i+1}) \mathbb{1}_{X^{(i)}}(x_i). \end{aligned}$$

Denoting

$$\begin{aligned} g_i(x_i, x_{i+1}) &:= f_i(x_i, x_{i+1}) \mathbb{1}_{X^{(i)}}(x_i), \\ \tilde{g}_i(x_i, x_{i+1}) &:= \mathbf{E}(f_i \mid \mathcal{P}' \otimes \mathcal{P}')(x_i, x_{i+1}) \mathbb{1}_{X^{(i)}}(x_i), \end{aligned}$$

Lemma 37 implies that

$$\|g_i - \tilde{g}_i\|_{S^{r_i}} \leq \|f_i - \mathbf{E}(f_i \mid \mathcal{P}' \otimes \mathcal{P}')\|_{S^{r_i}} \leq \frac{\varepsilon}{kC^{k-1}}.$$

We can thus apply our counting lemma, Lemma 51, for the spaces $X_i = X$ and the functions g_i, \tilde{g}_i , to obtain

$$\left| \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k g_i(x_i, x_{i+1}) - \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \tilde{g}_i(x_i, x_{i+1}) \right| \leq \varepsilon.$$

The point of splitting X into k parts and inserting the indicator functions $\mathbb{1}_{X^{(i)}}(x_i)$ was to force the product $\prod_{i=1}^k g_i(x_i, x_{i+1})$ to vanish when x_1, \dots, x_k are not pairwise distinct; it follows that

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k g_i(x_i, x_{i+1}) \leq \mathbf{E}_{x_1, \dots, x_k} \mathbb{1}_{x_1, \dots, x_k \text{ distinct}} \prod_{i=1}^k f_i(x_i, x_{i+1}),$$

and thus putting things together, we have

$$\begin{aligned}
\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i \mid \mathcal{P}' \otimes \mathcal{P}')(x_i, x_{i+1}) \mathbb{1}_{R^c}(x_i) &\ll_k \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \tilde{g}_i(x_i, x_{i+1}) \\
&\leq \varepsilon + \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k g_i(x_i, x_{i+1}) \\
&\leq \varepsilon + \mathbf{E}_{x_1, \dots, x_k} \mathbb{1}_{x_1, \dots, x_k \text{ distinct}} \prod_{i=1}^k f_i(x_i, x_{i+1}).
\end{aligned}$$

But the functions $\mathbf{E}(f_i \mid \mathcal{P}' \otimes \mathcal{P}')$ and $\mathbf{E}(f_i \mid \mathcal{P} \otimes \mathcal{P})$ are equal inside $R^c \times R^c$, so the same conclusion holds for \mathcal{P} instead of \mathcal{P}' . \square

5 Transference-free results

When counting paths or even-length cycles, Proposition 55 already has some interesting direct applications, with no transference step (and thus no dense removal lemma or tower-exponential bounds) required. The main relevant result here is the following.

Theorem 57 (Expected number of paths and even-length cycles). *Fix $\varepsilon, C > 0$ and integers $j \geq 1$, $k > \ell \geq 2$. Then for all $n \geq n_0(\varepsilon, C, j, k, \ell)$ and all $p > 0$, any n -vertex graph G satisfying*

$$\text{Hom}(P_1, G) \geq \varepsilon p n^2 \quad \text{and} \quad \text{Hom}(C_{2\ell}, G) \leq C p^{2\ell} n^{2\ell}$$

also satisfies

$$\text{Inj}(P_j, G) \asymp p^j n^{j+1} \quad \text{and} \quad \text{Inj}(C_{2k}, G) \asymp p^{2k} n^{2k}.$$

Remark 58. To clarify our notation, the implied constants may depend on the fixed parameters, but not on n and p . It is worth noting that a random (Erdős–Rényi) graph with $\asymp p n^2$ edges has an expected number of $\asymp p^j n^{j+1}$ P_j 's and $\asymp p^{2k} n^{2k}$ C_{2k} 's, and the conclusion of Theorem 57 confirms these predictions under relatively mild assumptions. What makes this result nontrivial is the fact that we are counting *injective* copies of P_j and C_{2k} ; if we counted homomorphic copies instead, the result would be a relatively straightforward corollary of the fact that paths and even-length cycles are bipartite graphs with Sidorenko's property [43].

We work our way to proving Theorem 57, starting with a couple of propositions for weighted graphs.

Proposition 59 (Weighted graphs with few paths). *Let $k \in \mathbb{Z}_{\geq 1}$, $q \in [1, \infty)$, $\varepsilon, C > 0$, and X be a large enough finite set (in terms of ε, k, q, C), equipped with uniform probability. If $f : X \times X \rightarrow [0, \infty)$ is symmetric (i.e., $f(x, y) = f(y, x), \forall x, y \in X$), $\|f\|_{S^q} \leq C$, and*

$$\mathbf{E}_{x_1, \dots, x_{k+1}} \mathbb{1}_{x_1, \dots, x_{k+1} \text{ distinct}} \prod_{i=1}^k f(x_i, x_{i+1}) \leq \varepsilon,$$

then $\|f\|_{L^1} \ll_k \varepsilon^{1/k} + C\varepsilon$.

Proof. Apply part (iii) of Proposition 55 with $k+1$ in place of k , $q_1 = \dots = q_k = q$ and $q_{k+1} = 1$, and the functions $f_1 = \dots = f_k = f$ and $f_{k+1} = \mathbb{1}_{X_k \times X_1}$ (recall that $\|f_{k+1}\|_{S^1} = 1$ by Lemma 37). This produces a partition \mathcal{P} of X and a \mathcal{P} -measurable subset $R \subset X$, such that $\mathbf{P}(R) \leq \varepsilon^2$ and

$$\mathbf{E}_{x_1, \dots, x_{k+1}} \prod_{i=1}^k \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{P})(x_i, x_{i+1}) \mathbb{1}_{R^c}(x_i) \ll_k \varepsilon.$$

Letting $g := \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{P}) \mathbb{1}_{R^c \times R^c}$, we thus have $\mathbf{E}_{x_1, \dots, x_{k+1}} \prod_{i=1}^k g(x_i, x_{i+1}) \ll_k \varepsilon$. But since g is symmetric and nonnegative, an older result of Blakley and Roy [7] implies that

$$\mathbf{E}_{x_1, \dots, x_{k+1}} \prod_{i=1}^k g(x_i, x_{i+1}) \geq (\mathbf{E}_{x_1, x_2} g(x_1, x_2))^k.$$

(Blakley and Roy state their result in the form $\langle u, G^k u \rangle \geq \langle u, Gu \rangle^k$, for any nonnegative symmetric $n \times n$ matrix G , and any $u \in \mathbb{R}^n$ with $\|u\|_2 = 1$; we can recover the inequality above by letting $n := |X|$, $G := \text{Mat}(g)$, and $u_x := \sqrt{1/|X|}$ for all $x \in X$. This can also be interpreted as the statement that paths satisfy Sidorenko's conjecture [43].)

We thus get $\|g\|_{L^1} \ll_k \varepsilon^{1/k}$. But $R^c \times R^c$ is $(\mathcal{P} \otimes \mathcal{P})$ -measurable, so

$$\begin{aligned} \varepsilon^{1/k} &\gg_k \|g\|_{L^1} = \|f \mathbb{1}_{R^c \times R^c}\|_{L^1} \\ &\geq \|f\|_{L^1} - 2\|f \mathbb{1}_{R \times X}\|_{L^1}. \end{aligned}$$

Recalling from (4) that

$$\|f \mathbb{1}_{R \times X}\|_{L^1} \leq \|f\|_{S^\infty} \|\mathbb{1}_R\|_{L^2} \|\mathbb{1}_X\|_{L^2} \leq \|f\|_{S^q} \sqrt{\mathbf{P}(R)} \leq C\varepsilon,$$

we reach the desired conclusion, $\|f\|_{L^1} \ll_k \varepsilon^{1/k} + C\varepsilon$. \square

Proposition 60 (Weighted graphs with few even cycles). *Let k be an even positive integer, $q \in [1, k)$, $\varepsilon, C > 0$, and X be a large enough finite set (in terms of ε, k, q), equipped with uniform probability. If $f : X \times X \rightarrow [0, \infty)$ is symmetric, $\|f\|_{S^q} \leq C$, and (working with indices modulo k)*

$$\mathbf{E}_{x_1, \dots, x_k} \mathbb{1}_{x_1, \dots, x_k \text{ distinct}} \prod_{i \in \mathbb{Z}/k\mathbb{Z}} f(x_i, x_{i+1}) \leq \varepsilon,$$

then $\|f\|_{L^1} \ll_k \varepsilon^{1/k} + C\varepsilon$.

Proof. Apply part (iii) of Proposition 55 for the functions $f_i = f$, to obtain a partition \mathcal{P} of X and a \mathcal{P} -measurable subset $R \subset X$, such that $\mathbf{P}(R) \leq \varepsilon^2$ and

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i \in \mathbb{Z}/k\mathbb{Z}} \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{P})(x_i, x_{i+1}) \mathbb{1}_{R^c}(x_i) \ll_k \varepsilon.$$

Let $g := \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{P}) \mathbb{1}_{R^c \times R^c}$. Since k is an even integer and g is real symmetric, (3) implies that the left-hand side above is precisely $\|g\|_{S^k}^k$. Since g is nonnegative, we further have

$$\|g\|_{L^1} = \|g\|_{\square} \leq \|g\|_{S^\infty} \leq \|g\|_{S^k} \ll_k \varepsilon^{1/k}.$$

(The bound $\|g\|_{L^1} \leq \|g\|_{S^k}$ is essentially the statement that even-length cycles satisfy Sidorenko's conjecture [43].)

But $R^c \times R^c$ is $(\mathcal{P} \otimes \mathcal{P})$ -measurable, so $\|g\|_{L^1} = \|f \mathbb{1}_{R^c \times R^c}\|_{L^1}$. As in the previous proof, this further implies that $\|f\|_{L^1} \ll_k \varepsilon^{1/k} + C\varepsilon$ since $\mathbf{P}(R) \leq \varepsilon^2$. \square

We take a final intermediate step to proving Theorem 57 through the following result, which resembles a removal lemma.

Proposition 61. *For any integers $j \geq 1$ and $k > \ell \geq 2$, and any $\varepsilon, C > 0$, there exist $n_0, \delta > 0$ such that the following holds. For any $n \geq n_0$ and $p > 0$, if an n -vertex graph G with $\text{Hom}(C_{2\ell}, G) \leq Cp^{2\ell}n^{2\ell}$ has*

$$\text{Inj}(P_j, G) \leq \delta p^j n^{j+1} \quad \text{or} \quad \text{Inj}(C_{2k}, G) \leq \delta p^{2k} n^{2k},$$

then G has fewer than $\varepsilon p n^2$ edges.

Proof. Let $q := 2\ell$, $X := V(G)$ (equipped with uniform probability) and $f := \mathbb{1}_G/p : X \times X \rightarrow [0, \infty)$. Then by (5),

$$\|f\|_{S^q}^q = \|\mathbb{1}_G/p\|_{S^{2\ell}}^{2\ell} = \frac{\text{Hom}(C_{2\ell}, G)}{p^{2\ell} n^{2\ell}} \leq C,$$

so $\|f\|_{S^q} \leq C^{1/q}$, and we aim to show that $\|f\|_{L^1} \leq \varepsilon$. But our second assumption about G states that

$$\mathbf{E}_{x_1, \dots, x_{j+1}} \mathbb{1}_{x_1, \dots, x_{j+1} \text{ distinct}} \prod_{i=1}^j f(x_i, x_{i+1}) \leq \delta,$$

or

$$\mathbf{E}_{x_1, \dots, x_{2k}} \mathbb{1}_{x_1, \dots, x_{2k} \text{ distinct}} \prod_{i \in \mathbb{Z}/2k\mathbb{Z}} f(x_i, x_{i+1}) \leq \delta,$$

for a value of δ that we have yet to choose. Propositions 59 and 60 now imply that for all large enough n ,

$$\|f\|_{L^1} \ll_j \delta^{1/j} + C^{1/q}\delta \quad \text{or} \quad \|f\|_{L^1} \ll_k \delta^{1/2k} + C^{1/q}\delta,$$

and we can choose δ small enough so that the upper bounds above are less than ε . \square

Proof of Theorem 57. Apply Proposition 61 for the same values of $j, k, \ell, \varepsilon, C$, to obtain some $n_0, \delta > 0$ such that the following contrapositive statement holds. For any $n \geq n_0$ and $p > 0$, and any n -vertex graph G with $\geq \varepsilon p n^2$ edges and $\text{Hom}(C_{2\ell}, G) \leq Cp^{2\ell}n^{2\ell}$, one has

$$\text{Inj}(P_j, G) > \delta p^j n^{j+1} \quad \text{and} \quad \text{Inj}(C_{2k}, G) > \delta p^{2k} n^{2k}.$$

We wish to find constants α, β (depending on $j, k, \ell, \varepsilon, C$) such that

$$\alpha p^j n^{j+1} \leq \text{Inj}(P_j, G) \leq \beta p^j n^{j+1} \quad \text{and} \quad \alpha p^{2k} n^{2k} \leq \text{Inj}(C_{2k}, G) \leq \beta p^{2k} n^{2k},$$

and we can already take $\alpha := 1/\delta$ for the desired lower bounds. For the upper bounds, first note that by (5), the condition $\text{Hom}(C_{2\ell}, G) \leq Cp^{2\ell} n^{2\ell}$ means $\|\mathbb{1}_G/p\|_{S^{2\ell}} \leq C^{1/2\ell}$. Since $\|\mathbb{1}_G/p\|_{S^{2k}} \leq \|\mathbb{1}_G/p\|_{S^{2\ell}}$, we get

$$\text{Inj}(C_{2k}, G) \leq \text{Hom}(C_{2k}, G) = n^{2k} \|\mathbb{1}_G\|_{S^{2k}}^{2k} \leq C^{k/\ell} p^{2k} n^{2k}.$$

Similarly, after fixing an ordering of the vertices of G (and equipping $V(G)$ with the uniform probability measure), one has

$$\text{Inj}(P_j, G) \leq \text{Hom}(P_j, G) = \sum_{v_1, \dots, v_{j+1} \in V(G)} \prod_{i=1}^k \mathbb{1}_G(v_i, v_{i+1}) = n^{j+1} \langle u, \text{Mat}(\mathbb{1}_G)^j u \rangle,$$

where $u \in \mathbb{R}^n$ is the constant vector with each entry equal to $1/\sqrt{n}$ (and thus $\|u\|_2 = 1$). We now have

$$\langle u, \text{Mat}(\mathbb{1}_G)^j u \rangle \leq \|\text{Mat}(\mathbb{1}_G)^j\|_{S^\infty} = \|\text{Mat}(\mathbb{1}_G)\|_{S^\infty}^j = \|\mathbb{1}_G\|_{S^\infty}^j,$$

since the singular values of a self-adjoint matrix are just the absolute values of its eigenvalues (thus the singular values of $\text{Mat}(\mathbb{1}_G)^j$ are the j th powers of the singular values of $\text{Mat}(\mathbb{1}_G)$). Putting things together, we get

$$\text{Inj}(P_j, G) \leq n^{j+1} \|\mathbb{1}_G\|_{S^\infty}^j \leq n^{j+1} \|\mathbb{1}_G\|_{S^{2\ell}}^j \leq C^{j/2\ell} p^j n^{j+1}.$$

So we can take $\beta := \max(C^{k/\ell}, C^{j/2\ell})$, and our proof is complete. \square

Next, we prove the first two parts Corollary 15, and leave the last part (which requires a transference step) to Section 6.

Theorem 62 (Rephrasing of parts **(i)**, **(ii)** of Corollary 15). *For any integers $j \geq 1$ and $k \geq 3$, and any $\varepsilon, C > 0$, there exist $n_0, \delta > 0$ such that the following holds. For any $n \geq n_0$ and $p \geq n^{-1/2}$, if an n -vertex graph G with $\text{Inj}(C_4, G) \leq Cp^4 n^4$ has*

$$\text{Inj}(P_j, G) \leq \delta p^j n^{j+1} \quad \text{or} \quad \text{Inj}(C_{2k}, G) \leq \delta p^{2k} n^{2k},$$

then G has fewer than $\varepsilon p n^2$ edges.

Proof. This follows directly from Proposition 61 by taking $\ell = 2$, provided that we can use the additional lower bound $p \geq n^{-1/2}$ to upgrade the upper bound $\text{Inj}(C_4, G) \leq Cp^4 n^4$ to

$$\text{Hom}(C_4, G) \leq C' p^4 n^4,$$

for another constant C' (depending on C). But we already did this in the proof of Corollary 53, using (21). \square

6 The transference step

In this section, we complete the transference step and prove our main removal lemmas for k -cycles (and k -paths) in weighted sparse graphs. We will “transfer” to the following weighted version of the k -cycle removal lemma for dense graphs (i.e., assuming L^∞ bounds); this is a fairly classical result, which can be proven along the same lines as [14, Theorem 4.1] (see also [51]).

Theorem 63 (Dense cycle removal). *For any $k \in \mathbb{Z}_{\geq 3}$ and $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let $(X_i, \Sigma_i, \mathbf{P}_i)$ be any probability spaces and $f_i : X_i \times X_{i+1} \rightarrow [0, 1]$ be measurable functions, for $i \in \mathbb{Z}/k\mathbb{Z}$. If*

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) \leq \delta, \quad (26)$$

then there exist measurable subsets $E_i \subset X_i \times X_{i+1}$ with $\|f_i \mathbb{1}_{E_i^c}\|_{L^1} \leq \varepsilon$ for each i , such that

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbb{1}_{E_i}(x_i, x_{i+1}) = 0.$$

Remark 64. The conclusion of this theorem (and similar ones) is sometimes [51] phrased with the vanishing of $\prod_{i=1}^k (f_i \mathbb{1}_{E_i})(x_i, x_{i+1})$ rather than $\prod_{i=1}^k \mathbb{1}_{E_i}(x_i, x_{i+1})$. To recover our statement from this one, we can simply intersect E_i with the support of f_i .

Remark 65. Thinking of f_i as the weight functions of a k -partite graph with few k -cycles, $\text{supp}(f_i) \setminus E_i$ represent the (small) sets of edges that one can remove in order to obtain a k -cycle free graph. If one is interested in general weighted graphs, one may take $X_1 = \dots = X_k = X$, $f_1 = \dots = f_k = f$,

$$E := \bigcap_{i=1}^k E_i,$$

and bound $\|f \mathbb{1}_{E^c}\|_{L^1}$ by the triangle inequality. If one is further interested in undirected graphs (so f is also symmetric), one can always modify E by considering the minimal symmetric set containing it. Of course, one issue with passing to non- k -partite graphs this way is that due to vertex collisions, (26) now assumes a bound for homomorphic copies of k -cycles rather than proper k -cycles, which matters for $k > 3$ (unlike in Theorem 1). Luckily, this issue can be ammended through the last part of Proposition 55 (the proof of which used Lemma 54).

Theorem 66 (Sparse cycle removal). *For any $k \in \mathbb{Z}_{\geq 3}$ and $q_1, \dots, q_k \in [1, \infty)$ with $\sum_i \frac{1}{q_i} > 1$, and any $\varepsilon, K, C > 0$, there exist $\delta, \tau > 0$ such that the following holds. Let (X_i, \mathbf{P}_i) be finite probability spaces and $f_i : X_i \times X_{i+1} \rightarrow [0, \infty)$, for $i \in \mathbb{Z}/k\mathbb{Z}$. Assume that for each i ,*

(i) (Bounded Schatten norm). *One has $\|f_i\|_{S^{q_i}} \leq C$, and*

(ii) (Dense pairs condition with large parts). *For any partitions $\mathcal{P}_i, \mathcal{P}_{i+1}$ of X_i, X_{i+1} into parts with probabilities $\geq \tau$, one has*

$$\|f_i \mathbb{1}_{\mathbf{E}(f_i | \mathcal{P}_i \otimes \mathcal{P}_{i+1}) > K} \|_{L^1} \leq \frac{\varepsilon}{4}. \quad (27)$$

If additionally

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) \leq \delta, \quad (28)$$

then there exist subsets $E_i \subset X_i \times X_{i+1}$ with $\|f_i \mathbb{1}_{E_i}\|_{L^1} \leq \varepsilon$ for each i , such that

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbb{1}_{E_i}(x_i, x_{i+1}) = 0.$$

Moreover, if all $(X_i, \mathbf{P}_i) = (X, \mathbf{P})$ are equal and \mathbf{P} is the uniform distribution, and $|X|$ is sufficiently large in terms of $k, q_1, \dots, q_k, \varepsilon, K, C$, then in the dense pairs condition one can assume that $\mathcal{P}_i = \mathcal{P}_{i+1}$, and one can replace (28) with

$$\mathbf{E}_{x_1, \dots, x_k} \mathbb{1}_{x_1, \dots, x_k \text{ distinct}} \prod_{i=1}^k f_i(x_i, x_{i+1}) \leq \delta.$$

If additionally all $f_i = f$ are equal, then one can also replace (27) with

$$\|f \mathbb{1}_{\mathbf{E}(f | \mathcal{P} \otimes \mathcal{P}) > K} \mathbb{1}_{\text{diag}(\mathcal{P})^c}\|_{L^1} \leq \frac{\varepsilon}{8}.$$

Remark 67. We recall that for a partition \mathcal{P} of a set X , we denoted $\text{diag}(\mathcal{P}) := \bigcup_{A \in \mathcal{P}} (A \times A) \subset X \times X$ in Section 2.2. The fact that we can ignore the contribution of the set $\text{diag}(\mathcal{P})$ to the average in (27) when all f_i 's are equal will be relevant for the last part of Corollary 15.

Remark 68. One can obtain a sparse k -path removal lemma similar to Theorem 66, by using the appropriate dense removal lemma, and applying our counting lemma (Lemma 51) with $f_1 = \tilde{f}_1 = \mathbb{1}_{X_1 \times X_2}$ and $q_1 = 1$; the condition $\sum_i \frac{1}{q_i} > 1$ can be discarded in that case. The resulting removal lemma is superseded by Proposition 59 when applied to a single, undirected graph (which is why we leave it to the reader), but it has new content for k -partite and/or directed graphs.

Proof. Denote $q := (q_1, \dots, q_k)$; we will choose a parameter $\delta = \delta(\varepsilon, K, k, C) > 0$ later in the proof. For now, we fix δ and apply part (ii) of Proposition 55 with δ in place of ε , to obtain a parameter $\tau = \tau(\delta, k, q, C) \in (0, 1)$ and partitions \mathcal{P}_i of X_i whose parts have probabilities $\geq \tau$ (except possibly for one part $R_i \in \mathcal{P}_i$ with $\mathbf{P}_i(R_i) \leq \delta^2$), such that

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f_i | \mathcal{P}_i \otimes \mathcal{P}_{i+1})(x_i, x_{i+1}) \mathbb{1}_{R_i^c}(x_i) \leq \delta + \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}).$$

By (28), the right-hand side is at most 2δ . Now consider, for each i , the $(\mathcal{P}_i \otimes \mathcal{P}_{i+1})$ -measurable functions

$$\tilde{f}_i := \mathbf{E}(f_i \mid \mathcal{P}_i \otimes \mathcal{P}_{i+1}) \quad \text{and} \quad g_i := \frac{\tilde{f}_i \mathbb{1}_{\tilde{f}_i \leq 4K}}{4K} \mathbb{1}_{R_i^c \times R_{i+1}^c}.$$

It follows that for some constant $c_{K,k} > 0$ depending on K and k , we have

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k g_i(x_i, x_{i+1}) \leq c_{K,k} \delta. \quad (29)$$

Note that $0 \leq g_i \leq 1$ everywhere, so we can apply the dense removal lemma (Theorem 63) for the probability spaces $(X_i, \Sigma(\mathcal{P}_i), \mathbf{P}_i)$ and the functions g_i . We finally pick $\delta = \delta(\varepsilon, K, k)$ small enough, as given by Theorem 63, to find $(\mathcal{P}_i \otimes \mathcal{P}_{i+1})$ -measurable subsets $E_i \subset X_i \times X_{i+1}$ such that

$$\prod_{i=1}^k \mathbb{1}_{E_i}(x_i, x_{i+1}) = 0 \quad \text{and} \quad \|g_i \mathbb{1}_{E_i^c}\|_{L^1} \leq \frac{\varepsilon}{16K},$$

for all x_1, \dots, x_k , and for all i respectively. We also make δ smaller if necessary, in terms of ε and C , to guarantee that

$$\delta \|f_i\|_{S^\infty} \leq \frac{\varepsilon}{4}.$$

(This is possible by the assumed Schatten norm bounds, recalling that $\|f_i\|_{S^\infty} \leq \|f_i\|_{S^{q_i}} \leq C$.)

It remains to show that $\|f_i \mathbb{1}_{E_i^c}\|_{L^1} \leq \varepsilon$. To this end, we split

$$\|f_i \mathbb{1}_{E_i^c}\|_{L^1} \leq \|f_i \mathbb{1}_{(R_i^c \times R_{i+1}^c)^c}\|_{L^1} + \|f_i \mathbb{1}_{R_i^c \times R_{i+1}^c} \mathbb{1}_{E_i^c}\|_{L^1}. \quad (30)$$

We use (4) to bound the first term:

$$\begin{aligned} \|f_i \mathbb{1}_{(R_i^c \times R_{i+1}^c)^c}\|_{L^1} &\leq \|f_i \mathbb{1}_{R_i \times X_{i+1}}\|_{L^1} + \|f_i \mathbb{1}_{X_i \times R_{i+1}}\|_{L^1} \\ &\leq \|f_i\|_{S^\infty} (\|\mathbb{1}_{R_i}\|_{L^2} + \|\mathbb{1}_{R_{i+1}}\|_{L^2}) \\ &\leq \|f_i\|_{S^\infty} (\sqrt{\mathbf{P}_i(R_i)} + \sqrt{\mathbf{P}_{i+1}(R_{i+1})}) \leq 2\|f_i\|_{S^\infty} \delta \leq \frac{\varepsilon}{2}. \end{aligned} \quad (31)$$

To handle the second term in (30), note that since $\{\tilde{f}_i \leq K\}$, E_i and $R_i^c \times R_{i+1}^c$ are $(\mathcal{P}_i \otimes \mathcal{P}_{i+1})$ -measurable sets, we have

$$\begin{aligned} \|f_i \mathbb{1}_{R_i^c \times R_{i+1}^c} \mathbb{1}_{E_i^c}\|_{L^1} &\leq \|f_i \mathbb{1}_{\tilde{f}_i > 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c}\|_{L^1} + \|f_i \mathbb{1}_{\tilde{f}_i \leq 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c} \mathbb{1}_{E_i^c}\|_{L^1} \\ &= \|f_i \mathbb{1}_{\tilde{f}_i > 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c}\|_{L^1} + \|\tilde{f}_i \mathbb{1}_{\tilde{f}_i \leq 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c} \mathbb{1}_{E_i^c}\|_{L^1} \\ &= \|f_i \mathbb{1}_{\tilde{f}_i > 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c}\|_{L^1} + \|4K g_i \mathbb{1}_{E_i^c}\|_{L^1} \\ &\leq \|f_i \mathbb{1}_{\tilde{f}_i > 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c}\|_{L^1} + \frac{\varepsilon}{4}. \end{aligned} \quad (32)$$

We are left to handle the term $\|\tilde{f}_i \mathbb{1}_{\tilde{f}_i > 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c}\|_{L^1}$, and we intend to use the dense pairs condition, recalling that $\tilde{f}_i = \mathbf{E}(f_i \mid \mathcal{P}_i \otimes \mathcal{P}_{i+1})$. However, the partitions \mathcal{P}_j (for $1 \leq j \leq k$) are not directly suitable for this condition, since one part of \mathcal{P}_j (which is R_j) may have probability $< \tau$. Hence if $\mathbf{P}_j(R_j) < \tau$, we define \mathcal{P}'_j by combining R_j with an arbitrary other part B_j of \mathcal{P}_j (which has $\mathbf{P}_j(B_j) \geq \tau$); otherwise we let $\mathcal{P}'_j := \mathcal{P}_j$. Denoting $f'_i := \mathbf{E}(f_i \mid \mathcal{P}'_i \otimes \mathcal{P}'_{i+1})$, we now claim that

$$\tilde{f}_i \mathbb{1}_{R_i^c \times R_{i+1}^c} \leq 4f'_i. \quad (33)$$

Both sides are $(\mathcal{P}_i \otimes \mathcal{P}_{i+1})$ -measurable, so we must check the inequality in every part $A_i \times A_{i+1}$ of $\mathcal{P}_i \otimes \mathcal{P}_{i+1}$. If $A_i = R_i$ or $A_{i+1} = R_{i+1}$, then the left-hand side vanishes. If $\mathbf{P}_i(R_i) < \tau$ and $A_i = B_i$, then we have

$$\begin{aligned} \mathbf{E}[f_i \mid B_i \times A_{i+1}] &= \frac{\mathbf{E}[f_i \mathbb{1}_{B_i \times A_{i+1}}]}{\mathbf{P}_i(B_i)\mathbf{P}_{i+1}(A_{i+1})} \\ &\leq 2 \frac{\mathbf{E}[f_i \mathbb{1}_{(B_i \cup R_i) \times A_{i+1}}]}{\mathbf{P}_i(B_i \cup R_i)\mathbf{P}_{i+1}(A_{i+1})} = 2\mathbf{E}[f_i \mid (B_i \cup R_i) \times A_{i+1}], \end{aligned}$$

since $\mathbf{P}_i(B_i) \geq \tau > \mathbf{P}_i(R_i)$ implies $2\mathbf{P}_i(B_i) > \mathbf{P}_i(B_i \cup R_i)$. One can repeat this reasoning if $\mathbf{P}_{i+1}(R_{i+1}) < \tau$ and $A_{i+1} = B_{i+1}$, showing that $\tilde{f}_i \mathbb{1}_{R_i^c \times R_{i+1}^c} \leq 4f'_i$ whenever $A_i \in \{B_i, R_i\}$ or $A_{i+1} \in \{B_{i+1}, R_{i+1}\}$. In all other cases, the values of \tilde{f}_i and f'_i are equal. Thus (33) holds, implying that $f'_i > K$ whenever $\tilde{f}_i > 4K$ and $\mathbb{1}_{R_i^c \times R_{i+1}^c} \neq 0$. Hence we have the pointwise bound

$$f_i \mathbb{1}_{\tilde{f}_i > 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c} \leq f_i \mathbb{1}_{f'_i > K}$$

everywhere, which implies

$$\|f_i \mathbb{1}_{\tilde{f}_i > 4K} \mathbb{1}_{R_i^c \times R_{i+1}^c}\|_{L^1} \leq \|f_i \mathbb{1}_{f'_i > K}\|_{L^1} \leq \varepsilon/4, \quad (34)$$

by our dense pairs condition (note that each part of \mathcal{P}'_j has probability $\geq \tau$).

Combining (30) to (32) and (34) completes our proof in the general case. The statement about equal probability spaces $(X_i, \mathbf{P}_i) = (X, \mathbf{P})$ (where \mathbf{P} is the uniform probability distribution on X) follows analogously, using part (iii) of Proposition 55 instead of part (ii) (in this case we have all $\mathcal{P}_i = \mathcal{P}$ and all $R_i = R$). It remains to prove the additional claim that (27) can be replaced with

$$\|f \mathbb{1}_{\mathbf{E}(f \mid \mathcal{P}' \otimes \mathcal{P}') > K} \mathbb{1}_{\text{diag}(\mathcal{P}')^c}\|_{L^1} \leq \frac{\varepsilon}{8}, \quad (35)$$

provided that we also have $f_1 = \dots = f_k = f$; here \mathcal{P}' is any partition of X into parts with probabilities $\geq \tau$. To this end, we will show that by choosing a smaller value of δ , (35) actually self-improves to (27) (in the form $\|f \mathbb{1}_{\mathbf{E}(f \mid \mathcal{P}' \otimes \mathcal{P}') > K}\|_{L^1} \leq \frac{\varepsilon}{4}$), for the specific partitions \mathcal{P}' that were used in the preceding proof (to obtain (34)). More precisely, our application of Proposition 55.(iii) produced a partition \mathcal{P} of X and a part $R \in \mathcal{P}$ with

$\mathbf{P}(R) \leq \delta^2$, such that

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{P})(x_i, x_{i+1}) \mathbb{1}_{R^c}(x_i) \ll_k \delta + \mathbf{E}_{x_1, \dots, x_k} \mathbb{1}_{x_1, \dots, x_k \text{ distinct}} \prod_{i=1}^k f(x_i, x_{i+1}) \leq 2\delta, \quad (36)$$

and the dense pairs condition was applied for a partition \mathcal{P}' , obtained by possibly combining R with another part of \mathcal{P} . Now fixing some part $A \in \mathcal{P}$ with $A \neq R$ (and thus $A \cap R = \emptyset$), and restricting the implied summation in (36) to those $x_1, \dots, x_k \in A$, we get

$$\begin{aligned} 2\delta &\gg_k \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbf{E}[f \mid A \times A] \mathbb{1}_A(x_i) = \mathbf{E}[f \mid A \times A]^k \mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbb{1}_A(x_i) \\ &= \mathbf{E}[f \mid A \times A]^k \mathbf{P}(A)^k. \end{aligned}$$

Summing over A yields

$$\sum_{\substack{A \in \mathcal{P} \\ A \neq R}} \mathbf{E}[f \mid A \times A] \mathbf{P}(A)^2 \ll_k \sum_{A \in \mathcal{P}} (2\delta)^{1/k} \mathbf{P}(A) = (2\delta)^{1/k},$$

and using the bound $\|f\|_{S^\infty} \leq \|f\|_{S^{q_1}} \leq C$ (coupled with (4)), we further obtain

$$\begin{aligned} \|f \mathbb{1}_{\text{diag}(\mathcal{P}')} \|_{L^1} &= \sum_{A \in \mathcal{P}'} \mathbf{E}[f \mathbb{1}_{A \times A}] \\ &\leq \mathbf{E}[f \mathbb{1}_{R \times X}] + \mathbf{E}[f \mathbb{1}_{X \times R}] + \sum_{A \in \mathcal{P}'} \mathbf{E}[f \mathbb{1}_{(A \setminus R) \times (A \setminus R)}] \\ &\leq 2\|f\|_{S^\infty} \sqrt{\mathbf{P}(R)} + \sum_{\substack{A \in \mathcal{P} \\ A \neq R}} \mathbf{E}[f \mathbb{1}_{A \times A}] \\ &\leq 2C\delta + \sum_{\substack{A \in \mathcal{P} \\ A \neq R}} \mathbf{E}[f \mid A \times A] \mathbf{P}(A)^2 \ll_k C\delta + (2\delta)^{1/k}. \end{aligned}$$

By making δ smaller (if necessary) in terms of ε, k, C , we can thus guarantee that $\|f \mathbb{1}_{\text{diag}(\mathcal{P}')} \|_{L^1} \leq \varepsilon/8$. Combining this with (35) recovers the original condition in (27) (for this specific partition \mathcal{P}' , in the form $\|f \mathbb{1}_{\mathbf{E}(f \mid \mathcal{P}' \otimes \mathcal{P}') > K} \|_{L^1} \leq \varepsilon/4$), and this completes our proof. \square

To make sense of the “dense pairs condition with large parts” and to compare our result with prior work, let us also state the immediate corollaries of Theorem 66 for (unweighted) graphs.

Corollary 69 (Cycle removal in multiple sparse graphs). *For any $k \in \mathbb{Z}_{\geq 3}$, $q_1, \dots, q_k \in [1, \infty)$ with $\sum_i \frac{1}{q_i} > 1$, and $\varepsilon, K, C > 0$, there exist $n_0, \delta, \tau > 0$ such that the following holds. Let $n \geq n_0$, $p_1, \dots, p_k > 0$, and G_1, \dots, G_k be n -vertex graphs with the same vertex set V , such that for each i ,*

(i) (Bounded Schatten norm). *Equipping V with the uniform probability distribution, one has $\|\mathbb{1}_{G_i}\|_{S^{q_i}}^{q_i} \leq Cp_i^{q_i}$;*

(ii) (Dense pairs condition with large parts). *For any partition of V into parts V_1, V_2, \dots with $\geq \tau n$ elements each, a total of at most $\varepsilon p_i n^2/4$ edges of G_i lie between pairs $(V_j, V_{j'})$ with*

$$e_{G_i}(V_j, V_{j'}) \geq Kp_i |V_j| |V_{j'}|.$$

(Here we allow pairs with $j = j'$.) If additionally

$$\sum_{x_1, \dots, x_k \text{ distinct}} \mathbb{1}_{G_1}(x_1, x_2) \cdots \mathbb{1}_{G_k}(x_k, x_1) \leq \delta p_1 \cdots p_k n^k,$$

then there exist subgraphs G'_i of G_i with $|E(G_i) \setminus E(G'_i)| \leq \varepsilon p_i n^2$ for each i , such that

$$\sum_{x_1, \dots, x_k} \mathbb{1}_{G'_1}(x_1, x_2) \cdots \mathbb{1}_{G'_k}(x_k, x_1) = 0.$$

Remark 70. Recall that if q_i is an even integer, then $\|\mathbb{1}_{G_i}\|_{S^{q_i}}^{q_i} = n^{-q_i} \text{Hom}(C_{q_i}, G_i)$, so condition (i) becomes a bound on G_i 's homomorphic even cycle counts.

Proof. This is just Theorem 66 applied to $X = V$, $f_i = \mathbb{1}_{G_i}/p_i$, $q_i = 2\ell_i$, 2ε instead of ε , and $C^{1/\max q_i}$ instead of C . \square

We can similarly deduce our main removal lemma from Section 1, concerning a single graph:

Proof of Theorem 9. This is just (the last paragraph of) Theorem 66 applied to $X = V(G)$, $f_i = \mathbb{1}_G/p$ and $q_i = 2k$ for $i \in \{1, \dots, 2k+1\}$, $2\varepsilon/(2k+1)$ instead of ε , and $C^{1/(2k+1)}$ instead of C . Unlike in Corollary 69, it now suffices to consider the ‘off-diagonal’ dense pairs (V_i, V_j) with $i \neq j$. Indeed, the set $\text{diag}(\mathcal{P})$ from Theorem 66 corresponds to the ‘diagonal’ pairs (V_i, V_i) , which can be ignored in the dense pairs condition since $f_1 = \cdots = f_{2k+1}$.

Also, note that Theorem 66 produces a graph which contains no homomorphic copies of C_{2k+1} , and such a graph cannot contain any cycle $C_{2\ell+1}$ with $1 \leq \ell \leq k$. \square

7 Applications

7.1 Graphs with mildly-pseudorandom majorants

Theorem 66 also implies a *relative* removal result, assuming the existence of a majorant $\nu \geq f$ satisfying the mild pseudorandomness assumption $\|\nu - 1\|_{\square} = o(1)$. The fact that the partitions from the dense pairs condition in Theorem 66 consist of parts of lower-bounded probability will be critical here.

Corollary 71 (Relative cycle removal). *For any $k \in \mathbb{Z}_{\geq 3}$ and $q_1, \dots, q_k \in [1, \infty)$ with $\sum_i \frac{1}{q_i} > 1$, and any $\varepsilon, K, C > 0$, there exist $\delta, \tau > 0$ such that the following holds. Let (X_i, \mathbf{P}_i) be finite probability spaces and $f_i, \nu_i : X_i \times X_{i+1} \rightarrow [0, \infty)$, for $i \in \mathbb{Z}/k\mathbb{Z}$. Assume that for each i , one has*

- (i) (Bounded Schatten norm). $\|f_i\|_{S^{q_i}} \leq C$,
- (ii) (Mildly-pseudorandom majorant). $f_i \leq K\nu_i$ and $\|\nu_i - 1\|_{\square} \leq \tau$.

If additionally

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) \leq \delta, \quad (37)$$

then there exist subsets $E_i \subset X_i \times X_{i+1}$ with $\|f_i \mathbb{1}_{E_i}\|_{L^1} \leq \varepsilon$ for each i , such that

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \mathbb{1}_{E_i}(x_i, x_{i+1}) = 0.$$

Moreover, if all $(X_i, \mathbf{P}_i) = (X, \mathbf{P})$ are equal and \mathbf{P} is the uniform probability distribution, and $|X|$ is sufficiently large in terms of $k, q_1, \dots, q_k, \varepsilon, K, C$, then instead of (37) one can require that $\mathbf{E}_{x_1, \dots, x_k} \mathbb{1}_{x_1, \dots, x_k \text{ distinct}} \prod_{i=1}^k f_i(x_i, x_{i+1}) \leq \delta$.

Remark 72. The pseudorandomness assumption in Corollary 71 is fairly general for the following reasons:

- (1). Firstly, for any function $\nu : X \times Y \rightarrow [0, \infty)$, (1) and (4) imply that

$$\|\nu - 1\|_{\square} = \sup_{\substack{A \subset X \\ B \subset Y}} |\mathbf{E}[(\nu - 1) \mathbb{1}_{A \times B}]| \leq \sup_{\substack{A \subset X \\ B \subset Y}} \frac{|\mathbf{E}[(\nu - 1) \mathbb{1}_{A \times B}]|}{\sqrt{\mathbf{P}_X(A) \mathbf{P}_Y(B)}} \leq \|\nu - 1\|_{S^{\infty}}.$$

When $X = Y = V$ is equipped with the uniform probability distribution on $|V| = n$ vertices, the second-to-last quantity above is $o(1)$ iff for all subsets $A, B \subset V$,

$$\left| \sum_{x \in A} \sum_{y \in B} \nu(x, y) - |A||B| \right| \leq o(n) \sqrt{|A||B|}.$$

Using terminology from [16], this is precisely the statement that ν is $(1, o(n))$ -jumbled, or equivalently, that $p\nu$ is $(p, o(pn))$ -jumbled for any $p > 0$. Therefore, our assumption that $\|\nu - 1\|_{\square} = o(1)$ is strictly weaker than the $(p, o(pn))$ -jumbledness of $p\nu$, which is weaker than $\|\nu - 1\|_{S^{\infty}} = o(1)$, and significantly weaker than $\|\nu - 1\|_{S^4} = o(1)$.

- (2). Secondly, the conditions $\|\nu_i - 1\|_{S^4} = o(1)$ are in turn weaker than the linear forms conditions of [17, 48], due to (3). Importantly, our assumptions about ν_i are separated, in the sense that each condition concerns only one of the majorants ν_i ; by

contrast, a linear forms-type condition would also include (among many others) an estimate of the form

$$\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k \nu_i(x_i, x_{i+1}) = 1 + o(1).$$

Of course, the linear forms conditions do not include a bounded-Schatten-norm condition for f_i , which is an advantage if $q_i < 4$, and particularly when $k = 3$; see point (4) below. But when $k = 3$, our result can still recover Roth's theorem in the primes, by choosing $q_i \in (2, 3)$ and using the *discrete majorant property* from Green's original proof [29, Lemma 6.6].

- (3). Thirdly, even the *dense model* theorem in [17] (which is the more flexible part of the transference process therein) assumes that $\|\nu - 1\|_{\square} = o(1)$, so this may be the weakest meaningful pseudorandomness condition that one can assume in practice.
- (4). Finally, provided that $q_i \geq 4$, one can let $r_i := \lfloor q_i/2 \rfloor$ and note that

$$\|f_i\|_{S^{q_i}} \leq \|f_i\|_{S^{2r_i}} \leq \|\nu_i\|_{S^{2r_i}},$$

by (3). Hence the bounded-Schatten-norm condition can be replaced by a pseudorandomness condition $\|\nu_i\|_{S^{2r_i}} \ll 1$. In particular, it suffices to assume that $\|\nu_i - 1\|_{S^{2r_i}} = o(1)$, which also supersedes the condition $\|\nu_i - 1\|_{\square} = o(1)$.

Proof of Corollary 71. Apply Theorem 66 with $2K$ in place of K and $\sqrt{\tau}$ in place of τ . This produces some $\varepsilon, \tau > 0$ such that our desired conclusion holds, assuming the following condition:

$$\text{For any } A_j \subset X_j \text{ with } \mathbf{P}_j(A_j) \geq \sqrt{\tau}, \text{ one has } \mathbf{E}[f_i \mid A_i \times A_{i+1}] \leq 2K. \quad (38)$$

Indeed, if (38) holds, then for any partitions $\mathcal{P}_i, \mathcal{P}_{i+1}$ of X_i, X_{i+1} into parts of probabilities $\geq \sqrt{\tau}$, we will have $\mathbf{E}(f_i \mid \mathcal{P}_i \otimes \mathcal{P}_{i+1}) \leq 2K$ everywhere, making the dense pairs condition in Theorem 66 trivial. It thus remains to verify (38); given such A_j , we can write

$$\begin{aligned} \mathbf{E}[f_i \mid A_i \times A_{i+1}] &\leq K \mathbf{E}[\nu_i \mid A_i \times A_{i+1}] \\ &= K + K \mathbf{E}[\nu_i - 1 \mid A_i \times A_{i+1}] \\ &\leq K + K \frac{\mathbf{E}[(\nu_i - 1) \mathbb{1}_{A_i \times A_{i+1}}]}{\mathbf{P}(A_i) \mathbf{P}(A_{i+1})} \leq K + K \frac{\|\nu_i - 1\|_{\square}}{\tau} \leq 2K \end{aligned}$$

for each i , which completes our proof. \square

For the remainder of this subsection, we deduce Corollaries 12 and 23 from Corollary 71.

Corollary 73 (Rephrasing of Corollary 12). *For any $k \in \mathbb{Z}_{\geq 2}$ and $\varepsilon, K, C > 0$, there exist $n_0, \delta, \tau > 0$ such that the following holds. Let $n \geq n_0$, and G be an n -vertex graph with $\text{Inj}(C_{2k+1}, G) \leq \delta p^{2k+1} n^{2k+1}$ and $\mathbb{1}_G/p \leq K\nu$ for some $\nu : V(G) \times V(G) \rightarrow [0, \infty)$ with $\|\nu\|_{S^{2k}} \leq C$ and $\|\nu - 1\|_{\square} \leq \tau$. Then one can remove at most $\varepsilon p n^2$ edges from G to obtain a graph G' with $\text{Hom}(C_{2k+1}, G') = 0$.*

Proof. This follows from (the last paragraph of) Corollary 71 applied to $X = V(G)$, $2k+1$ in place of k , $2\varepsilon/(2k+1)$ in place of ε , $f_i = \mathbb{1}_G/p$, $q_i = 2k$, and KC in place of C . Note that the condition $\|\nu\|_{S^{2k}} \leq C$ implies $\|f_i\|_{S^{2k}} \leq K\|\nu_i\|_{S^{2k}} \leq KC$, due to (3). \square

Corollary 74 (Extension of Corollary 23). *For any $k \in \mathbb{Z}_{\geq 3}$, $q \in [1, k)$, and $K, C > 0$, there exist $\delta, \tau > 0$ such that the following holds. Let a_1, \dots, a_k be nonzero integers with $a_1 + \dots + a_k = 0$, $n \in \mathbb{Z}_{\geq 1}$ with $\gcd(n, a_1 \cdots a_k) = 1$, G be an abelian group of order n , and $f, \nu : G \rightarrow [0, \infty)$ satisfy*

(i) (Not-too-large Fourier coefficients). $\|\widehat{f}\|_q \leq C\widehat{f}(0)$, and

(ii) (Mildly-pseudorandom majorant). $f \leq K\frac{\widehat{f}(0)}{n}\nu$ and $\|\nu - 1\|_{\square} \leq \tau$.

Then one has

$$\sum_{\substack{g_1, \dots, g_k \in G \\ a_1 g_1 + \dots + a_k g_k = 0}} f(g_1) \cdots f(g_k) \geq \delta \frac{\widehat{f}(0)^k}{n}.$$

Proof. Apply Corollary 71 for $q_i = k$ and $1/(2k)$ in place of ε to obtain values of $\delta, \tau > 0$, and let f be as in the corollary's statement. The conclusion is trivial if f is the identically zero function, so we may assume without loss of generality that $\widehat{f}(0) \neq 0$. Moreover, we may scale f by a constant to further assume that $\widehat{f}(0) = n$ (in other words, f has average 1).

Let $b_1, \dots, b_k \in \mathbb{Z} \setminus \{0\}$ be inverses modulo n for a_1, \dots, a_k , $X_1 = \dots = X_k := G$ (equipped with the uniform probability measure), and $f_i, \nu_i : X_i \times X_{i+1} \rightarrow [0, \infty)$ be given by

$$f_i(x_i, x_{i+1}) := f(b_i(x_i - x_{i+1})), \quad \nu_i(x_i, x_{i+1}) := \nu(b_i(x_i - x_{i+1})),$$

for $i \in \mathbb{Z}/k\mathbb{Z}$. By Lemma 40 and assumption (i), we have $\|f_i\|_{S^q} = \frac{1}{n}\|\widehat{f}\|_q \leq C$. Moreover, $\|\nu_i - 1\|_{\square} = \|\nu - 1\|_{\square}$ since the latter was defined as the cut norm of the function $(x, y) \mapsto (\nu - 1)(x + y)$, and the maps $x \mapsto \pm b_i x$ are automorphisms of G . Thus by assumption (ii), we have $f_i \leq K\nu_i$ and $\|\nu_i - 1\|_{\square} \leq \tau$.

Now assume towards a contradiction that $\sum_{a_1 g_1 + \dots + a_k g_k = 0} f(g_1) \cdots f(g_k) < \delta n^{k-1}$, which translates to $\mathbf{E}_{x_1, \dots, x_k} \prod_{i=1}^k f_i(x_i, x_{i+1}) < \delta$. Then we may apply the conclusion of Corollary 71 to get subsets $E_i \subset X_i \times X_{i+1}$ with $\|f_i \mathbb{1}_{E_i^c}\|_{L^1} \leq 1/(2k)$ for each i , such that $\prod_{i=1}^k \mathbb{1}_{E_i}(x_i, x_{i+1}) = 0$ for any x_1, \dots, x_k . In particular, this implies that $\sum_{i=1}^k \mathbb{1}_{E_i^c}(x_i, x_{i+1}) \geq 1$ always.

To obtain a contradiction, we define $T \subset G^k$ by

$$T := \{(x_1, \dots, x_k) \in X_1 \times \dots \times X_k : b_1(x_1 - x_2) = b_2(x_2 - x_3) = \dots = b_k(x_k - x_1)\}.$$

Note that any pair $(x_i, x_{i+1}) \in X_i \times X_{i+1}$ uniquely determines a k -tuple $(x_1, \dots, x_k) \in T$ containing it; the uniqueness follows easily from the definition of T , while the existence follows from the fact that $a_1 + \dots + a_k = 0$. Indeed, denoting $g_i = b_i(x_i - x_{i+1})$, we have

$\sum_{i=1}^k a_i g_i = 0$; thus if we assume that $k - 1$ of the g_i 's are equal, then all of them must be equal. Moreover, for $(x_1, \dots, x_k) \in T$, all of the values $f_i(x_i, x_{i+1}) = f(b_i(x_i - x_{i+1}))$ are equal. Hence we may write

$$\begin{aligned}
\|f_1\|_{L^1} &= \frac{1}{n^2} \sum_{(x_1, \dots, x_k) \in T} f_1(x_1, x_2) \\
&\leq \frac{1}{n^2} \sum_{(x_1, \dots, x_k) \in T} f_1(x_1, x_2) \sum_{i=1}^k \mathbb{1}_{E_i^c}(x_i, x_{i+1}) \\
&= \frac{1}{n^2} \sum_{i=1}^k \sum_{(x_1, \dots, x_k) \in T} f_i(x_i, x_{i+1}) \mathbb{1}_{E_i^c}(x_i, x_{i+1}) \\
&= \sum_{i=1}^k \frac{1}{n^2} \sum_{(x_i, x_{i+1}) \in X_i \times X_{i+1}} f_i \mathbb{1}_{E_i^c}(x_i, x_{i+1}) = \sum_{i=1}^k \|f_i \mathbb{1}_{E_i^c}\|_{L^1},
\end{aligned}$$

which implies that $1 \leq k(1/2k) = 1/2$. Thus $\sum_{a_1 g_1 + \dots + a_k g_k = 0} f(g_1) \cdots f(g_k) \geq \delta n^{k-1}$, as we wanted. \square

Proof of Corollary 23. Take $q = 2\ell$ and $f = \mathbb{1}_S$ in Corollary 74, and note that $\|\widehat{\mathbb{1}_S}\|_{2\ell} = n\|\mathbb{1}_S\|_{S^{2\ell}} \leq \varepsilon^{-1}|S| \cdot \|\nu\|_{S^{2\ell}} \leq \varepsilon^{-1}C|S|$ by (i) and (ii). \square

7.2 Graphs with few 4-cycles

Following [14], we first state a lemma relating the dense pairs condition to counts of 4-cycles.

Lemma 75 (Dense pairs with large parts I). *Let $\tau, q > 0$, and G be an n -vertex graph with $T := \text{Inj}(C_4, G)$. Consider a partition of $V(G)$ into parts V_1, V_2, \dots with at least τn elements each. Then G has at most*

$$O\left(\frac{n}{q} + \frac{1}{q\tau^2} + \frac{T^{1/4}n}{\tau}\right) \tag{39}$$

edges lying between pairs $(V_j, V_{j'})$ with $e_G(V_j, V_{j'}) \geq q|V_j||V_{j'}|$. Moreover, if one only counts the ‘off-diagonal’ pairs with $j \neq j'$, then one can replace T with

$$\sup_{G \supset H \text{ bipartite}} \text{Inj}(C_4, H).$$

Proof. This is a slight simplification of the proof of [14, Lemma 2.4], since in our case we know that $|V_j| \geq \tau n$ for each j .

Let W_j be the union of all $V_{j'}$'s with $e_G(V_j, V_{j'}) \geq q|V_j||V_{j'}|$; in particular, $e_G(V_j, W_j) \geq q|V_j||W_j|$. Let us consider the bipartite graph H_j with vertex sets V_j and W_j^0 , where W_j^0 is a copy of W_j (so that $W_j^0 \cap V_j = \emptyset$), and edges induced by the edges of G . While H_j is not necessarily isomorphic to a subgraph of G (since we may have $V_j \subset W_j$, in which case

the vertices in V_j have copies in W_j^0), it is true that every 4-cycle in H_j induces a 4-cycle in G (the fact that G is a simple graph causes the right pairs of vertices to be distinct in $V(G)$). Moreover, each 4-cycle in G corresponds to at most two 4-cycles in H_j , and thus $\text{Inj}(C_4, H_j) \ll \text{Inj}(C_4, G)$. This line of reasoning is implicit in [14, Lemma 2.4].

Now if $e(V_j, W_j) < 4|V_j||W_j|^{1/2} + 4|W_j|$, then we have

$$e(V_j, W_j) \leq \frac{e(V_j, W_j)^2}{q|V_j||W_j|} \ll \frac{|V_j|^2|W_j| + |W_j|^2}{q|V_j||W_j|} = \frac{|V_j|}{q} + \frac{|W_j|}{q|V_j|} \leq \frac{|V_j|}{q} + \frac{1}{q\tau}.$$

Otherwise, provided that W_j is nonempty, [14, Lemma 2.3] implies that

$$e(V_j, W_j)^4|V_j|^{-2}|W_j|^{-2} \ll \text{Inj}(C_4, H_j) \ll \text{Inj}(C_4, G) = T.$$

Thus

$$e(V_j, W_j) \ll T^{1/4}|V_j|^{1/2}|W_j|^{1/2} \leq T^{1/4}n.$$

Thus in both cases, we have

$$e(V_j, W_j) \ll \frac{|V_j|}{q} + \frac{1}{q\tau} + T^{1/4}n.$$

Summing over all j (which gives a total of at most $n/(\tau n) = 1/\tau$ terms), we recover the desired bound in (39).

If we only count the ‘off-diagonal’ pairs $(V_j, V_{j'})$ with $j \neq j'$, then we may repeat the same reasoning by replacing W_j with $W_j \setminus V_j$. The difference is that H_j is now isomorphic to a subgraph of G , since there are no repeated vertices in V_j and W_j ; thus we have

$$\text{Inj}(C_4, H_j) \leq \sup_{G \supset H \text{ bipartite}} \text{Inj}(C_4, H),$$

and we may use the right-hand side as the parameter T . □

We now state a further corollary of Corollary 69 for graphs with few 4-cycles, which should be compared with the result for k -partite graphs in [14, Theorem 1.6] (with $k = 5$).

Corollary 76 (Cycle removal in multiple graphs with few 4-cycles). *For any $k \in \mathbb{Z}_{\geq 5}$ and any $\varepsilon, C > 0$, there exist $n_0, \delta > 0$ such that the following holds. Let $n \geq n_0$, $p_1, \dots, p_k \geq n^{-1/2}$, and G_1, \dots, G_k be n -vertex graphs with the same vertex set V , such that for each i ,*

$$\text{Inj}(C_4, G_i) \leq \delta p_i^4 n^4. \quad (40)$$

If additionally

$$\sum_{x_1, \dots, x_k \text{ distinct}} \mathbb{1}_{G_1}(x_1, x_2) \cdots \mathbb{1}_{G_k}(x_k, x_1) \leq \delta p_1 \cdots p_k n^k,$$

then there exist subgraphs G'_i of G_i with $|E(G_i) \setminus E(G'_i)| \leq \varepsilon p_i n^2$ for each i , such that

$$\sum_{x_1, \dots, x_k} \mathbb{1}_{G'_1}(x_1, x_2) \cdots \mathbb{1}_{G'_k}(x_k, x_1) = 0.$$

Proof. Fix $k \in \mathbb{Z}_{\geq 5}$ and $\varepsilon > 0$, and let $n_0, \delta, \tau > 0$ be obtained from Corollary 69 with $q_1 = \dots = q_k := 4$, K a large enough multiple of $1/\varepsilon$, and C a large enough absolute constant. Note that we may always decrease our choice of δ and increase our choice of n_0 (in terms of ε and k), without breaking the conclusion of Corollary 69; so for a start, let us take $\delta \leq 1$ without loss of generality.

Now let $n \geq n_0$, $p_1, \dots, p_k \geq n^{-1/2}$, and G_1, \dots, G_k be graphs with the same vertex set V (of size n), such that $\text{Inj}(C_4, G_i) \leq \delta p_i^4 n^4$. To complete our proof, it suffices to verify that each G_i satisfies the conditions in Corollary 69:

- (i). One has $\text{Hom}(C_4, G_i) \leq C(p_i n)^4$;
- (ii). G_i obeys the dense pairs condition with parts of sizes $\geq \tau n$.

To verify condition (i), note that (21) implies

$$\text{Hom}(C_4, G) \ll n^2 + \delta p_i^4 n^4 \ll p_i^4 n^4,$$

which is enough since we chose C to be a large enough absolute constant.

Concerning condition (ii), let $V = V_1 \cup V_2 \cup \dots$ be a partition into parts of sizes $\geq \tau n$. Applying Lemma 75 with $q := Kp_i$ and $T = \text{Inj}(C_4, G_i) \leq \delta p_i^4 n^4$, we find that G_i has at most

$$O\left(\frac{n}{Kp_i} + \frac{1}{Kp_i \tau^2} + \frac{\delta^{1/4} p_i n^2}{\tau}\right)$$

edges between pairs $(V_j, V_{j'})$ with $e_G(V_j, V_{j'}) \geq Kp_i |V_j| |V_{j'}|$. We would like this to be at most equal to $\varepsilon p_i n^2 / 4$, which can be achieved if:

- (1). K is a large enough multiple of $1/\varepsilon$ and $p_i^2 n \geq 1$ (both of which are true);
- (2). $1/\tau^2 \leq n$, which is true provided that we increase the value of n_0 accordingly;
- (3). $\delta^{1/4}/\tau$ is smaller than an absolute multiple of ε , which is true provided that we decrease the value of δ accordingly.

This completes our proof with (new) values of n_0 and δ depending only on ε and k (since K and τ also only depend on ε and k). \square

We can similarly deduce the last part of Corollary 15, which concerns a single graph G . In fact, we prove a slightly more general statement, where we assume that $\text{Hom}(C_{2k}, G) \ll p^{2k} n^{2k}$ instead of $\text{Inj}(C_4, G) \ll p^4 n^4$; the former follows from the latter due to the inequalities (21), $p \geq n^{-1/2}$, and $\|\mathbb{1}_G/p\|_{S^{2k}} \leq \|\mathbb{1}_G/p\|_{S^4}$.

Corollary 77 (Generalizing part (iii) of Corollary 15). *For any $k \in \mathbb{Z}_{\geq 2}$ and $\varepsilon, C > 0$, there exist $n_0, \delta > 0$ such that the following holds. For $n \geq n_0$ and $p \geq n^{-1/2}$, any n -vertex graph G with $\text{Hom}(C_{2k}, G) \leq Cp^{2k} n^{2k}$, $\sup_{G \supset H \text{ bipartite}} \text{Inj}(C_4, H) \leq \delta p^4 n^4$, and $\text{Inj}(C_{2k+1}, G) \leq \delta p^{2k+1} n^{2k+1}$ can be made $\{C_3, C_5, \dots, C_{2k+1}\}$ -free by removing at most $\varepsilon p n^2$ edges.*

Proof. We apply Theorem 9 with the given value of C , and repeat the same reasoning as in the previous proof to verify the dense pairs condition. The difference is that it now suffices to count the ‘off-diagonal’ pairs $(V_j, V_{j'})$ with $j \neq j'$, which allows us to use $T = \sup_{G \supset H \text{ bipartite}} \text{Inj}(C_4, H)$ in Lemma 75. \square

Finally, we prove Corollary 28, rewriting the rational slopes as fractions $s_i = a_i/b_i$ (where $(a_i, b_i) = (1, 0)$ if $s_i = \infty$).

Corollary 78 (Rephrasing of Corollary 28). *For any $\varepsilon > 0$, $k \in \mathbb{Z}_{\geq 5}$, and $(a_i, b_i) \in \mathbb{Z}^2$ such that $\gcd(a_i, b_i) = 1$ and $a_i b_{i+1} \neq a_{i+1} b_i$ for $i \in \mathbb{Z}/k\mathbb{Z}$, there exists $\delta > 0$ such that the following holds. Let $n \in \mathbb{Z}_{\geq 1}$ and $S \subset [n]^2$ satisfy*

(i). $|S| \geq \varepsilon n^{3/2}$, and

(ii). For each i , S contains $\leq \delta |S|^4/n^4$ solutions in distinct points $(x_1, y_1), \dots, (x_4, y_4)$ to

$$\begin{aligned} a_i(x_2 - x_1) &= b_i(y_2 - y_1), & a_{i+1}(x_3 - x_2) &= b_{i+1}(y_3 - y_2), \\ a_i(x_4 - x_3) &= b_i(y_4 - y_3), & a_{i+1}(x_1 - x_4) &= b_{i+1}(y_1 - y_4). \end{aligned}$$

Then S contains at least $\delta |S|^k/n^k$ solutions in $(x_1, y_1), \dots, (x_k, y_k)$ to the equations

$$a_i(x_{i+1} - x_i) = b_i(y_{i+1} - y_i), \quad i \in \mathbb{Z}/k\mathbb{Z}. \quad (41)$$

Proof. Let $A := k(2 \max_i(|a_i| + |b_i|) + 1)$. We choose $n_0, \delta_0 > 0$ as in Corollary 76, for the same value of k and with $\varepsilon/(4kA^2)$ in place of ε ; we set $\delta := \delta_0 / \max(\varepsilon^4, \varepsilon^k)$ here.

Let n, S be as in the corollary’s statement; we may assume without loss of generality that $n \geq n_0$, since the conclusion is trivial for $n \leq n_0$ provided that $\delta < n_0^{2-k}$ (due to the $|S|$ diagonal solutions $(x_1, y_1) = \dots = (x_k, y_k)$).

For $i \in \mathbb{Z}/k\mathbb{Z}$, we define the invertible 2×2 matrix

$$M_i := \begin{pmatrix} a_i & b_i \\ a_{i+1} & b_{i+1} \end{pmatrix}$$

and the set

$$V_i := [-(|a_i| + |b_i|)n, (|a_i| + |b_i|)n] \cap \mathbb{Z}.$$

Finally, we let V be a disjoint union of all the V_i ’s (so in particular $n_0 \leq n \leq |V| \leq An$), and define the (bipartite) graph G_i with vertex set V by

$$\mathbb{1}_{G_i}(u, v) := \mathbb{1}_{V_i \times V_{i+1}}(u, v) \mathbb{1}_S(M_i^{-1} \begin{pmatrix} u \\ v \end{pmatrix}),$$

where we identify pairs $(x, y) \in S$ with vectors $\begin{pmatrix} x \\ y \end{pmatrix}$. Note that for $(x_i, y_i), (x_{i+1}, y_{i+1}) \in S$, we have $a_i(x_{i+1} - x_i) = b_i(y_{i+1} - y_i)$ iff $a_i x_i + b_i y_i = a_{i+1} x_{i+1} + b_{i+1} y_{i+1} =: u_i$ for some $u_i \in V_i$. With this change of variables, the system of equations in (41) can be rewritten as

$$M_i \begin{pmatrix} x_{i+1} \\ y_{i+1} \end{pmatrix} = \begin{pmatrix} u_i \\ u_{i+1} \end{pmatrix}, \quad i \in \mathbb{Z}/k\mathbb{Z},$$

so that the number of solutions to (41) becomes $\sum_{u_1, \dots, u_k \in V} \prod_{i=1}^k \mathbb{1}_{G_i}(u_i, u_{i+1})$; we want to show that this is at least equal to $\delta|S|^k/n^k$. Suppose for the sake of contradiction that this is false, so

$$\sum_{u_1, \dots, u_k \in V} \prod_{i=1}^k \mathbb{1}_{G_i}(u_i, u_{i+1}) \leq \frac{\delta_0}{\varepsilon^k} \frac{|S|^k}{n^k} = \delta_0 p_1 \cdots p_k n^k \leq \delta_0 p_1 \cdots p_k |V|^k,$$

where $p_1 = \cdots = p_k := |S|/(\varepsilon n^2) \geq n^{-1/2}$ (by assumption (i)). Similarly, assumption (ii) implies that for each i ,

$$\sum_{\substack{u_1, u_2, u_3, u_4 \in V \\ u_1 \neq u_3, u_2 \neq u_4}} \mathbb{1}_{G_i}(u_1, u_2) \mathbb{1}_{G_i}(u_3, u_2) \mathbb{1}_{G_i}(u_1, u_4) \mathbb{1}_{G_i}(u_2, u_4) \leq \frac{\delta|S|^4}{n^4} \leq \frac{\delta_0}{\varepsilon^4} \frac{|S|^4}{n^4} \leq \delta_0 p_i^4 |V|^4,$$

which is to say, $\text{Inj}(C_4, G) \leq \delta_0 p_i^4 |V|^4$. By the conclusion of our initial application of Corollary 76, we find that there exist subsets $E_i \subset V_i \times V_{i+1}$ such that

$$\sum_{u_i, u_{i+1}} \mathbb{1}_{G_i} \mathbb{1}_{E_i^c}(u_i, u_{i+1}) \leq \frac{\varepsilon}{2kA^2} p_i |V|^2 \leq \frac{\varepsilon n^2}{2k} p_i = \frac{|S|}{2k}$$

for each i , and also $\sum_{u_1, \dots, u_k \in V} \prod_{i=1}^k \mathbb{1}_{E_i^c}(u_i, u_{i+1}) = 0$; in other words, $\sum_{i=1}^k \mathbb{1}_{E_i^c}(u_i, u_{i+1}) \geq 1$ everywhere. To obtain a contradiction, we consider the set

$$T := \{(u_1, \dots, u_k) \in V_1 \times \cdots \times V_k : M_1^{-1}(\frac{u_1}{u_2}) = \cdots = M_k^{-1}(\frac{u_k}{u_1}) \in S\},$$

which has natural bijections to S and to the edges of each G_i (through S). We conclude that

$$\begin{aligned} |S| = |T| &\leq \sum_{(u_1, \dots, u_k) \in T} \sum_{i=1}^k \mathbb{1}_{E_i^c}(u_i, u_{i+1}) \\ &= \sum_{i=1}^k \sum_{(u_1, \dots, u_k) \in T} \mathbb{1}_{E_i^c}(u_i, u_{i+1}) \\ &= \sum_{i=1}^k \sum_{(u_i, u_{i+1}) \in V_i \times V_{i+1}} \mathbb{1}_{G_i} \mathbb{1}_{E_i^c}(u_i, u_{i+1}) \leq k \frac{|S|}{2k} = \frac{|S|}{2}, \end{aligned}$$

a contradiction. Thus it must be the case that there are at least $\delta|S|^k/n^k$ solutions to (41), as we wanted. \square

7.3 Graphs free of quasi-smooth families

It remains to prove Corollaries 19 and 26; we start with a lemma similar to Lemma 75, but in the context of quasi-smooth families (recall Definition 17).

Lemma 79 (Dense pairs with large parts II). *Let $2 > \alpha > \beta \geq 1$ and \mathcal{F} be an (α, β) -quasi-smooth family of bipartite graphs. Also, let $\tau, q > 0$, G be an \mathcal{F} -free n -vertex graph, and $V_1 \cup V_2 \cup \dots$ be a partition $V(G)$ into parts with at least τn elements each. Then G has at most*

$$O_{\mathcal{F}} \left(\frac{n}{q^{(\alpha-1)/(2-\alpha)}} \left(1 + \frac{1}{n^{(\alpha-\beta)/(2-\alpha)} \tau^{\alpha/(2-\alpha)}} \right) \right)$$

edges lying between pairs $(V_j, V_{j'})$ with $e(V_j, V_{j'}) \geq q|V_j||V_{j'}|$ and $j \neq j'$. If \mathcal{F} only consists of complete bipartite graphs $K_{s,t}$, then one can also include the diagonal pairs (V_j, V_j) .

Proof. For each j , denote by W_j the union of all parts $V_{j'}$ with $e(V_j, V_{j'}) \geq q|V_j||V_{j'}|$ and $j \neq j'$; thus in particular $e(V_j, W_j) \geq q|V_j||W_j|$ and $V_j \cap W_j = \emptyset$. Let H_j be the bipartite subgraph of G with vertex sets V_j and W_j , and $e(V_j, W_j)$ edges. Then H_j is also \mathcal{F} -free, so by the the quasi-smoothness of \mathcal{F} we have

$$\begin{aligned} e(V_j, W_j) &\ll_{\mathcal{F}} \min(|V_j|, |W_j|) \max(|V_j|, |W_j|)^{\alpha-1} + \max(|V_j|, |W_j|)^{\beta} \\ &\leq |V_j||W_j|^{\alpha-1} + n^{\beta}, \end{aligned}$$

where we used that $\alpha - 1 \in (0, 1)$. If W_j is nonempty, this further implies that upon setting $\gamma := \frac{1}{2-\alpha}$,

$$\begin{aligned} e(V_j, W_j) &\ll_{\mathcal{F}} \frac{e(V_j, W_j)^{\gamma}}{q^{\gamma-1}|V_j|^{\gamma-1}|W_j|^{\gamma-1}} \\ &\ll \frac{|V_j|^{\gamma}|W_j|^{\gamma(\alpha-1)} + n^{\gamma\beta}}{q^{\gamma-1}|V_j|^{\gamma-1}|W_j|^{\gamma-1}} \leq \frac{|V_j|}{q^{\gamma-1}} + \frac{n^{\gamma\beta-2(\gamma-1)}}{(q\tau^2)^{\gamma-1}}. \end{aligned}$$

Summing over all (at most $1/\tau$) values of j and noting that $\gamma\beta - 2(\gamma - 1) = 1 - (\alpha - \beta)/(2 - \alpha)$, we recover the desired upper bound.

Finally, if \mathcal{F} only consists of complete bipartite graphs, then it is okay to include V_j into W_j (provided that $e(V_j, V_j) \geq q|V_j|^2$). Indeed, as in the proof of Lemma 75, we can let H_j have vertex sets V_j and W_j^0 instead, where W_j^0 is a copy of W_j that is disjoint with V_j . While H_j may not be isomorphic to a subgraph of G , it must remain $K_{s,t}$ -free for each $K_{s,t} \in \mathcal{F}$ (since the existence of such a $K_{s,t}$ would force all the involved vertices in V_j and W_j^0 to originate from distinct vertices in G , thus inducing a copy of $K_{s,t}$ in G). \square

Corollary 80 (Rephrasing of Corollary 19). *Fix $k \in \mathbb{Z}_{\geq 2}$, $2 > \alpha > \beta \geq 1$, and an (α, β) -quasi-smooth family \mathcal{F} . Then for any $\varepsilon, C > 0$, there exist $n_0, \delta > 0$ such that the following holds. For $n \geq n_0$ and $p \geq n^{\alpha-2}$, any \mathcal{F} -free n -vertex graph G with $\text{Hom}(C_{2k}, G) \leq Cp^{2k}n^{2k}$ and $\text{Inj}(C_{2k+1}, G) \leq \delta p^{2k+1}n^{2k+1}$ can be made $\{C_3, C_5, \dots, C_{2k+1}\}$ -free by removing at most εpn^2 edges.*

Proof. Fix ε, C and let $n_0, \delta, \tau > 0$ be obtained by applying Theorem 9 with the given value of C , and K a large enough constant (to be chosen later in terms of ε and α). Let $n \geq n_0$, $p \geq n^{\alpha-2}$, and G be an \mathcal{F} -free n -vertex graph with $\text{Hom}(C_{2k}, G) \leq Cp^{2k}n^{2k}$ and $\text{Inj}(C_{2k+1}, G) \leq \delta p^{2k+1}n^{2k+1}$.

Arguing as in the proof of Corollary 76, it suffices to verify the dense pairs condition in Theorem 9 (which only concerns ‘off-diagonal’ dense pairs $(V_j, V_{j'})$ with $j \neq j'$ and $e_G(V_j, V_{j'}) \geq Kp|V_j||V_{j'}|$). By Lemma 79 for $q = Kp$, G contains a total of at most

$$O\left(\frac{n}{(Kp)^{(\alpha-1)/(2-\alpha)}} \left(1 + \frac{1}{n^{(\alpha-\beta)/(2-\alpha)} \tau^{\alpha/(2-\alpha)}}\right)\right)$$

edges lying between such pairs, and we want this to be less than $\varepsilon pn^2/8$. This is true provided that n is large enough in terms of τ, α, β (which holds after increasing the value of n_0 accordingly), and that

$$\frac{n}{p^{(\alpha-1)/(2-\alpha)}} \leq pn^2, \quad (42)$$

since we chose K to be large enough in terms of ε and α . But (42) rearranges precisely to our assumption that $p \geq n^{\alpha-2}$, which completes our proof. \square

One can similarly deduce a corollary for multiple graphs, using Corollary 69 instead of Theorem 9; however, the necessity to consider diagonal dense pairs (V_j, V_j) restricts our result to families that consist only of complete bipartite graphs.

Corollary 81 (Cycle removal in multiple $K_{s,t}$ -free graphs). *For any $k \in \mathbb{Z}_{\geq 3}$, any $\varepsilon, C > 0$, and any $t_i \geq s_i \geq 1$ for $1 \leq i \leq k$, there exist $n_0, \delta > 0$ such that the following holds. Let $n \geq n_0$, $p_i \geq n^{-1/s_i}$, and G_1, \dots, G_k be n -vertex graphs with the same vertex set V , such that for each i ,*

- (i). *Equipping V with uniform probability, one has $\|\mathbb{1}_{G_i}\|_{S^{q_i}}^{q_i} \leq Cp_i^{q_i}$, and*
- (ii). *G_i is K_{s_i, t_i} -free.*

If additionally

$$\sum_{x_1, \dots, x_k \text{ distinct}} \mathbb{1}_{G_1}(x_1, x_2) \cdots \mathbb{1}_{G_k}(x_k, x_1) \leq \delta p_1 \cdots p_k n^k,$$

then there exist subgraphs G'_i of G_i with $|E(G_i) \setminus E(G'_i)| \leq \varepsilon p_i n^2$ for each i , such that

$$\sum_{x_1, \dots, x_k} \mathbb{1}_{G'_1}(x_1, x_2) \cdots \mathbb{1}_{G'_k}(x_k, x_1) = 0.$$

Proof. We follow the proof of Corollary 76, using Lemma 79 instead of Lemma 75 (and the given values of $q_i, p_i \geq n^{-1/s_i}$). The Schatten norm bounds are now given explicitly in our hypothesis, while the dense pairs condition follows from the same computation as in (42) for each i (recall that K_{s_i, t_i} is (α_i, β_i) -quasi-smooth [26] with $\alpha_i = 2 - 1/s_i$ and $\beta_i = 2 - 2/s_i$, so we have $p_i \geq n^{-1/s_i} = n^{\alpha_i-2}$). \square

Finally, we prove (a generalization of) Corollary 26, using a Sidon-type condition corresponding to $K_{s,t}$ -free graphs. We say that S is (s, t) -Sidon if and only if $S^{s \times t}$ contains no (s, t) -matrix M with distinct rows and distinct columns, such that

$$M_{i,i} + M_{j,j} = M_{i,j} + M_{j,i}$$

for all $1 \leq i \leq s$ and $1 \leq j \leq t$ (i.e., the rows of M are translations of each other by elements in S). This recovers classical Sidon sets when $s = t = 2$; note that (s, t) -Sidon sets are also (t, s) -Sidon and (s', t') -Sidon for $s' \geq s$ and $t' \geq t$. We expect that (s, t) -Sidon subsets of an abelian group of order n have size $\ll n^{1-1/\min(s,t)}$, so that the following corollary concerns the largest of these sets.

Corollary 82 (Extension of Corollary 26). *For any $\varepsilon, C > 0$, $s, t \in \mathbb{Z}_{\geq 2}$, $k \in \mathbb{Z}_{\geq 3}$ and $q \in [1, k]$, there exists $\delta > 0$ such that the following holds. Let a_1, \dots, a_k be nonzero integers with $a_1 + \dots + a_k = 0$, G be an abelian group of order n with $\gcd(n, a_1 \dots a_k) = 1$, and $S \subset G$ satisfy*

- (i). $\|\widehat{\mathbb{1}_S}\|_q \leq C|S|$,
- (ii). $|S| \geq \varepsilon n^{1-1/\min(s,t)}$,
- (iii). S is (s, t) -Sidon.

Then S contains $\geq \delta|S|^k/n$ solutions in x_1, \dots, x_k to the linear equation $a_1x_1 + \dots + a_kx_k = 0$.

Proof. We can assume without loss of generality that $t \geq s$ and that n is large enough in terms of $\varepsilon, C, s, t, k, q$. Indeed, for small values $n \leq n_0(\varepsilon, C, s, t, k, q)$, the conclusion is trivial once we take $\delta < n_0^{2-k}$, due to the $|S|$ diagonal solutions $x_1 = \dots = x_k$.

We follow the proof of Corollary 74 (with $f = n\mathbb{1}_S/|S|$ and a smaller choice of ε), using Corollary 81 instead of Corollary 71. We let V consist of k distinct copies V_1, \dots, V_k of the abelian group G of order n , and set $\mathbb{1}_{G_i}(x_i, x_{i+1}) := \mathbb{1}_{V_i \times V_{i+1}}(x_i, x_{i+1})\mathbb{1}_S(b_i(x_i - x_{i+1}))$, and $p := k^{-1/s}|S|/(\varepsilon n)$; note that we have $p \geq |V|^{-1/s}$ by the third assumption in our hypothesis. The conclusion that S contains at least $\delta|S|^k/n$ solutions to $a_1x_1 + \dots + a_kx_k$ is naturally the same as that of Corollary 23.

The only significant difference is that instead of a pseudorandom majorant condition, we now need to verify that each G_i is $K_{s,t}$ -free. Suppose for the sake of contradiction that this is false, so there exist distinct $x_1, \dots, x_s \in V_i = G$ and distinct $y_1, \dots, y_t \in V_{i+1} = G$ such that

$$\forall u \leq s, \forall v \leq t : \quad b_i(x_u - y_v) \in S.$$

But then, the matrix $M \in S^{s \times t}$ with entries given by $M_{u,v} := b_i(x_u - y_v)$ has distinct rows and columns (since $a_i b_i \equiv 1 \pmod{n}$, the x_u 's are distinct, and the y_v 's are distinct), and satisfies $M_{u,u} + M_{v,v} = M_{u,v} + M_{v,u}$ for each u and v . This contradicts the fact that S is (s, t) -Sidon, and completes our proof. \square

Remark 83. If S is an (s, t) -Sidon subset of $[n] := \{1, \dots, n\}$ instead of an abelian group, one can identify S with a subset of $\mathbb{Z}/N\mathbb{Z}$ with the same property, where $N \ll_{s,t,a_1, \dots, a_k} n$ can be chosen to be relatively prime with a_1, \dots, a_k . One can then combine Corollary 26 with an argument of Marcinkiewicz—Zygmund [54] (specifically, see [29, Lemma 6.5])

to obtain an analogous statement for (s, t) -Sidon subsets of $[n]$, where the condition on Fourier coefficients becomes continuous:

$$\int_{\mathbb{R}/\mathbb{Z}} \left| \frac{1}{|S|} \sum_{m \in S} e^{2\pi i \alpha m} \right|^q d\alpha \leq \frac{C}{n}.$$

Proof of Corollary 26. As noted in the remark above, it suffices to prove an analogous statement for subsets of abelian groups G of order n , where $\gcd(n, a_1 \cdots a_k) = 1$. Suppose that $S \subset G$ has $|S| \geq \varepsilon n^{1/2}$, and that for each $h \neq 0$, S contains at most C pairs of the form $(x, x + h)$. If t denotes any integer strictly larger than $\max(C, 1)$, this implies that $S^{2 \times t}$ contains no matrix of the form

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_t \\ x_1 + h & x_2 + h & \cdots & x_t + h \end{pmatrix},$$

where x_1, \dots, x_t are distinct. In other words, S is $(2, t)$ -Sidon.

The conclusion would now follow from Corollary 82, using $s = 2$ and $q = 4$, provided that we can show that

$$\|\widehat{\mathbb{1}_S}\|_4 \ll_{\varepsilon, C} |S|.$$

This is equivalent to a bound for the *additive energy* of S :

$$\# \{(a, b, c, d) \in S^4 : a - b = c - d\} \ll_{\varepsilon, C} \frac{|S|^4}{n}. \quad (43)$$

But we have

$$\begin{aligned} \# \{(a, b, c, d) \in S^4 : a - b = c - d\} &= \sum_{h \in G} \# \{(a, b, c, d) \in S^4 : a - b = c - d = h\} \\ &= \sum_{h \in G} (\# \{(a, b) \in S^2 : a - b = h\})^2. \end{aligned}$$

The contribution of $h = 0$ is exactly $|S|^2$. Moreover, by our second assumption, the contribution of each $h \neq 0$ is at most C^2 . Overall, we obtain

$$\# \{(a, b, c, d) \in S^4 : a - b = c - d\} \leq |S|^2 + C^2 n,$$

which is acceptable in (43) since $|S| \geq \varepsilon n^{1/2}$. This completes our proof. \square

A Spectral norms and entropy

In this section we prove Lemma 46. We begin with a quick lemma.

Lemma 84. *Consider a linear function $\lambda(z) = az + b$, with $a, b \in \mathbb{C}$.*

(i). *Given any (bounded) line segment $S \subset \mathbb{C}$, the maximum $\max_{z \in S} |\lambda(z)|$ is attained at one of the ends of S .*

(ii). Given any $z_0 \in \mathbb{C}$, there exists $z_1 \in \{0\} \cup \{2^k, i2^k, -2^k, -i2^k : k \in \mathbb{Z}_{\geq 0}\}$ such that $|z_1| \leq \max(1, 2\sqrt{2}|z_0|)$, and $|\lambda(z_1)| \geq |\lambda(z_0)|$.

Proof. Part (a) follows by writing $S = \{ct+d : t \in [0, 1]\}$ for some $c, d \in \mathbb{C}$, and expressing $|\lambda(ct+d)|^2$ as a convex function of t .

For part (b), draw a line segment through z_0 with slope in $\{\pm 1\}$, intersecting both semi-axes of z_0 's quadrant at w_1 and w_2 (if z_0 is purely real or imaginary, we may ignore this step). By part (a), we must have $|\lambda(z_0)| \leq |\lambda(w_i)|$ for some $i \in \{1, 2\}$. Then, draw a line segment between 0 and $2^k w_i / |w_i|$, where $k \geq 0$ is minimal such that $|w_i| \leq 2^k$, and apply part (a) again. \square

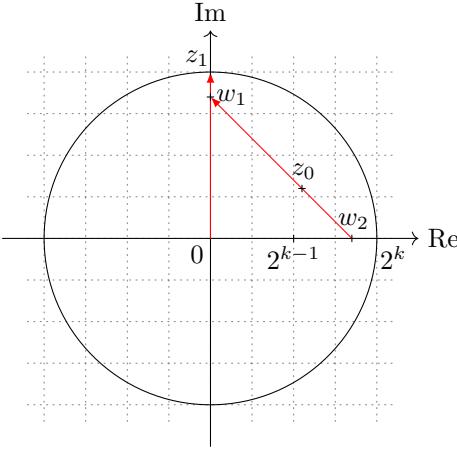


Figure 2: An example of finding z_1 starting from z_0 in Lemma 84.

We now recall the notation from (13),

$$\mathcal{F}(Z) := \{g : Z \rightarrow \{0\} \cup \{2^k, i2^k, -2^k, -i2^k : k \in \mathbb{Z}_{\geq 0}\} \mid \|g\|_{L^2} \leq 3\}.$$

Proof of Lemma 46. We first prove the entropy bound $\mathbf{H}(g) \ll 1$ for $g \in \mathcal{F}(Z)$. Denoting by $\mathbf{P}_Z(g = 0)$ the probability of the set $\{g = 0\} = g^{-1}(0)$, we have

$$\mathbf{H}(g) = \mathbf{P}_Z(g = 0) \log \frac{1}{\mathbf{P}_Z(g = 0)} + \sum_{k=0}^{\infty} \sum_{\omega^4=1} \mathbf{P}_Z(g = \omega 2^k) \log \frac{1}{\mathbf{P}_Z(g = \omega 2^k)},$$

and

$$9 \geq \|g\|_{L^2}^2 = \sum_{k=0}^{\infty} \sum_{\omega^4=1} 4^k \mathbf{P}_Z(g = \omega 2^k).$$

In particular, it follows that $\mathbf{P}_Z(g = \omega 2^k) \leq \frac{9}{4^k}$. Now consider the function $u(p) := p \log \frac{1}{p}$ on $[0, 1]$ (with $u(0) := 0$); this function is bounded, nonnegative, and strictly increasing near 0. Hence for sufficiently large k , we must have

$$u(\mathbf{P}_Z(g = \omega 2^k)) \leq u\left(\frac{9}{4^k}\right),$$

and thus

$$\mathbf{H}(g) \ll 1 + 4 \sum_{k=2}^{\infty} u \left(\frac{9}{4^k} \right) = 1 + 4 \sum_{k=2}^{\infty} \frac{9}{4^k} (k \log 4 - \log 9) \ll 1.$$

Now, to prove the upper bound on $\|f\|_{S^\infty}$, recall that

$$\|f\|_{S^\infty} = \sup_{\|g\|_{L^2} = \|h\|_{L^2} = 1} |\mathbf{E}_{x,y}[f(x, y)g(x)h(y)]|.$$

For fixed functions g, h with $\|g\|_{L^2} = \|h\|_{L^2} = 1$, the expectation $\mathbf{E}_{x,y}[f(x, y)g(x)h(y)]$ is linear in each value of g and h . Hence we can apply Lemma 84 to modify $g(x)$ for each $x \in X$, and $h(y)$ for each $y \in Y$, one at a time, while not decreasing the value of $|\mathbf{E}_{x,y}[f(x, y)g(x)h(y)]|$, until the resulting functions g' and h' only have values in $\{0\} \cup \{2^k, i2^k, -2^k, -i2^k : k \in \mathbb{Z}_{\geq 0}\}$. Then we have

$$|\mathbf{E}_{x,y}[f(x, y)g(x)h(y)]| \leq |\mathbf{E}_{x,y}[f(x, y)g'(x)h'(y)]|,$$

and

$$\|g'\|_{L^2}^2 = \mathbf{E}_x [|g'(x)|^2] \leq \mathbf{E}_x \left[1 + \left(2\sqrt{2}|g(x)| \right)^2 \right] = 1 + 8\|g\|_{L^2}^2 = 9,$$

so $g' \in \mathcal{F}(X)$, and similarly $h' \in \mathcal{F}(Y)$. The desired bound for $\|f\|_{S^\infty}$ follows after taking a supremum over g, h . \square

Remark 85. If in Lemma 46 we assumed that f takes real values (which is enough for all combinatorial applications), then in (4), it suffices to consider real-valued functions g, h . Consequently, one can ignore the possible values $\pm i2^k$ from Lemma 46 in this case.

B The “super-dense” removal lemma

In this section, we prove (a slight generalization of) Theorem 21. We will work with a finite probability space (X, \mathbf{P}) with full σ -algebra; all expectations involving variables $(x_u)_{u \in V(H)}$ are understood to sample each x_u from (X, \mathbf{P}) , independently of the other variables.

Theorem 86 (Weighted version of Theorem 21). *For any graph H , $q \in [1, 2)$, $C > 0$ and $\varepsilon \in (0, 1)$, there exists*

$$\delta \geq \exp(-O_{H,q,C}(\varepsilon^{-O_{H,q}(1)}))$$

such that the following holds. Let $f : X \times X \rightarrow [0, \infty)$ be a symmetric function (i.e., $f(x, y) = f(y, x)$) satisfying $\|f\|_{L^\infty} \leq 1$, $\|f\|_{S^q} \leq C$, and

$$\mathbf{E} \prod_{\{u,v\} \in E(H)} f(x_u, x_v) \leq \delta. \tag{44}$$

Then there is a symmetric set $E \subset X \times X$ such that $\|f \mathbb{1}_{E^c}\|_{L^1} \leq \varepsilon$ and

$$\prod_{\{u,v\} \in E(H)} \mathbb{1}_E(x_u, x_v) = 0$$

for any choice of the variables $(x_u)_{u \in V(H)}$.

Remark 87. One can remove the assumption $\|f\|_{L^\infty} \leq 1$ at the expense of a slightly worse bound. To recover Theorem 21, one can take $X = V(G)$ (equipped with uniform probability) and $f = \mathbb{1}_G$.

Proof. Note that $E(H)$ cannot be empty, since $\mathbf{E}[1] = 1 > \varepsilon$.

Let us apply our regularity lemma, Theorem 47, for some parameter $\tau \in (0, 1)$ to be chosen later in the proof (in place of ε), the function f , and $r = 2$. This yields a partition \mathcal{P} of X with

$$\mathbf{H}(\mathcal{P}) \leq \frac{c(q)}{\tau^{4/(2-q)}},$$

for some constant $c(q) > 0$, such that

$$\|f - \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{P})\|_{L^2} \leq \tau C.$$

For brevity, we write $f_1 := \mathbf{E}(f \mid \mathcal{P} \otimes \mathcal{P})$ and $f_2 := f - f_1$, so that f_1 is $(\mathcal{P} \otimes \mathcal{P})$ -measurable and $\|f_2\|_{L^2} \leq \tau C$. Note that this has the effect of a *strong* regularity lemma, but with no uniform component f_3 (and no tower-exponential dependencies).

We define $E^c \subset X \times X$ as the union of the following subsets $A \times B \in \mathcal{P} \otimes \mathcal{P}$, using three parameters α, β, γ to be chosen later:

- (i). Those where f_1 is small (more precisely, smaller than α);
- (ii). Those where $|f_2|$ is large on average (more precisely, where $\mathbf{E}[|f_2| \mid A \times B] \geq \beta$);
- (iii). Those with small probability (more precisely, with $\mathbf{P}(A) \leq \gamma$ or $\mathbf{P}(B) \leq \gamma$).

In other words, we defined E by removing the blocks of edges above from $X \times X$. Now suppose for the sake of contradiction that $\prod_{\{u,v\} \in E(H)} \mathbb{1}_E(x_u, x_v)$ is nonzero for some choice of $(x_u)_{u \in V(H)}$. Letting A_u be the part of \mathcal{P} containing x_u , we must have

$$(A_u, A_v) \in E, \quad \forall \{u, v\} \in E(H).$$

Assuming that (44) holds for some $\delta \in (0, 1)$ to be chosen later, we have

$$\mathbf{E} \left[\prod_{w \in V(H)} \mathbb{1}_{A_w}(x_w) \prod_{\{u,v\} \in E(H)} f(x_u, x_v) \right] \leq \delta,$$

and we may split $f = f_1 + f_2$ in each factor from the product above. In the resulting expansion, each term containing some instance of $f_2(x_u, x_v)$ is bounded, due to the L^∞ bound on f , by

$$\begin{aligned} \mathbf{E} \left[|f_2|(x_u, x_v) \prod_{w \in V(H)} \mathbb{1}_{A_w}(x_w) \right] &= \mathbf{E} \left[\mathbf{E}(|f_2| \mid \mathcal{P} \otimes \mathcal{P})(x_u, x_v) \prod_{w \in V(H)} \mathbb{1}_{A_w}(x_w) \right] \\ &= \mathbf{E} \left[\mathbf{E}[|f_2| \mid A_u \times A_v] \prod_{w \in V(H)} \mathbb{1}_{A_w}(x_w) \right] \\ &\leq \beta \prod_{w \in V(H)} \mathbf{P}(A_w). \end{aligned}$$

On the other hand, the term consisting only of f_1 's is lower-bounded, due to the removal of the type-(i) blocks, by

$$\alpha^{|E(H)|} \prod_{w \in V(H)} \mathbf{P}(A_w).$$

Putting the last three bounds together, we obtain $(\alpha^{|E(H)|} - \beta) \prod_{w \in V(H)} \mathbf{P}(A_w) \leq \delta$, and we now choose $\beta := \alpha^{|E(H)|}/2$ to conclude that

$$\delta \geq \frac{\alpha^{|E(H)|}}{2} \prod_{w \in V(H)} \mathbf{P}(A_w) \geq \frac{\alpha^{|E(H)|} \gamma^{|V(H)|}}{2}.$$

Finally, we set

$$\delta := \frac{\alpha^{|E(H)|} \gamma^{|V(H)|}}{4} \tag{45}$$

to obtain a contradiction (proving that $\prod_{\{u,v\} \in E(H)} \mathbb{1}_E(x_u, x_v)$ is zero everywhere). It remains to choose suitable values for τ, α, γ such that $\|f \mathbb{1}_{E^c}\|_{L^1} \leq \varepsilon$. By the definition of E^c ,

$$\|f \mathbb{1}_{E^c}\|_{L^1} \leq \mathbf{E}[f \mathbb{1}_{f_1 \leq \alpha}] + \mathbf{P}(\mathbf{E}(|f_2| \mid \mathcal{P} \otimes \mathcal{P}) \geq \beta) + \mathbf{P}(R_\gamma)^2,$$

where R_γ denotes the union of all sets $A \in \mathcal{P}$ with $\mathbf{P}(A) \leq \gamma$. But by the same reasoning as in (25), we have

$$\mathbf{P}(R_\gamma) \leq \frac{\mathbf{H}(\mathcal{P})}{\log(1/\gamma)} \leq \frac{c(q)}{\tau^{4/(2-q)} \log(1/\gamma)}.$$

At the same time, we have $\mathbf{E}[f \mathbb{1}_{f_1 \leq \alpha}] = \mathbf{E}[f_1 \mathbb{1}_{f_1 \leq \alpha}] \leq \alpha$ and

$$\mathbf{P}(\mathbf{E}(|f_2| \mid \mathcal{P} \otimes \mathcal{P}) \geq \beta) \leq \frac{\|f_2\|_{L^1}}{\beta} \leq \frac{2\|f_2\|_{L^2}}{\alpha^{|E(H)|}} \leq \frac{2\tau C}{\alpha^{|E(H)|}},$$

so it makes sense to pick

$$\alpha(\tau) := \tau^{1/2|E(H)|} \quad \text{and} \quad \gamma(\tau) := \exp\left(-\frac{1}{\tau^{1+4/(2-q)}}\right),$$

to conclude that

$$\|f \mathbb{1}_{E^c}\|_{L^1} \leq \tau^{1/2|E(H)|} + 2C\sqrt{\tau} + c(q)^2\tau^2.$$

To ensure that the right-hand side is at most equal to ε (and that $\tau \in (0, 1)$), it suffices to pick some

$$\tau = \left(\frac{\varepsilon}{O_{C,q}(1)} \right)^{2E(H)}.$$

Noting that both $\alpha(\tau)$ and $\gamma(\tau)$ are increasing functions of τ , this produces (due to (45)) a value of

$$\delta \gg \exp \left(- \left(\frac{\varepsilon}{O_{C,q}(1)} \right)^{-O_{H,q}(1)} \right)^{O_H(1)},$$

as we wanted. \square

C A sharp regularity lemma depending only on edge density

Recall that using Schatten norms in the upper bounds of regularity lemmas allowed us to work with sparse graphs G , with edge density $p = \mathbf{E}[\mathbb{1}_G] \searrow 0$, since the Schatten q -norm $\|\mathbb{1}_G/p\|_{S^q}$ is not a function of p (and may remain bounded as $p \searrow 0$ for certain families of graphs). We now ask for the best upper bound that one can use in a (weak) regularity lemma, which *only* depends on the edge density p . In fact, we will prove a more general *relative* regularity lemma, leading to a dense model theorem. We will apply our abstract energy optimization argument (Lemma 44) with a logarithmic version of energy, which may be of independent interest.

Definition 88 (Relative logarithmic energy). Let (X, \mathbf{P}) be a finite probability space (with full σ -algebra), \mathcal{P} be a partition of X , and $f, \nu : X \rightarrow [0, \infty)$ satisfy $f = 0$ whenever $\nu = 0$. Let us normalize f, ν such that $\mathbf{E}[f] = \mathbf{E}[\nu] = 1$. We define the *relative logarithmic energy* of \mathcal{P} with respect to f and ν by

$$\mathcal{E}_f^\nu(\mathcal{P}) := \mathbf{E} \left[f \log \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \right],$$

where we interpret ratios of $0/0$ as 0 . Note that the trivial partition has energy 0 .

Lemma 89 (Relative logarithmic energy increment). *Let $\mathcal{P}, \mathcal{P}'$ be partitions of a finite probability space (X, \mathbf{P}) such that \mathcal{P}' is finer than \mathcal{P} . Then one has*

$$\mathbf{E} \left| \mathbf{E}(f \mid \mathcal{P}') - \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \mathbf{E}(\nu \mid \mathcal{P}') \right| \leq \sqrt{2(\mathcal{E}_f^\nu(\mathcal{P}') - \mathcal{E}_f^\nu(\mathcal{P}))}.$$

In particular, this shows that relative logarithmic energy is nondecreasing and nonnegative.

Proof. This is a consequence of Pinsker's inequality in the form

$$\mathbf{E}|g - h| \leq \sqrt{2\mathbf{E} \left[g \log \frac{g}{h} \right]},$$

when $\mathbf{E}[g] = \mathbf{E}[h] = 1$ (and $g, h \geq 0$, with $g = 0$ whenever $h = 0$). \square

Theorem 90 (Weak relative regularity lemma). *Let $(X_1 \times X_2, \mathbf{P})$ be a finite probability space, $\varepsilon \in (0, 1)$, and $f, \nu : X_1 \times X_2 \rightarrow [0, \infty)$ be such that $f = 0$ whenever $\nu = 0$, and $\mathbf{E}[f] = \mathbf{E}[\nu] = 1$. Then there is a partition $\mathcal{P} = \mathcal{P}_1 \otimes \mathcal{P}_2$ of $X_1 \times X_2$ with $2^{O(1/\varepsilon^2)}$ parts, such that*

$$\left\| f - \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \nu \right\|_{\square} \leq \varepsilon \sqrt{\mathbf{E} \left[f \log \frac{f}{\nu} \right]}.$$

If $X_1 = X_2$, one can also take $\mathcal{P}_1 = \mathcal{P}_2$. In particular, if G is a finite graph with edge density $p = \mathbf{E}[\mathbb{1}_G]$ (allowing self-edges), one can find a partition \mathcal{P}_0 of $V(G)$ with $2^{O(1/\varepsilon^2)}$ parts such that

$$\|\mathbb{1}_G - \mathbf{E}(\mathbb{1}_G \mid \mathcal{P}_0 \otimes \mathcal{P}_0)\|_{\square} \leq \varepsilon p \sqrt{\log(1/p)},$$

where $\mathcal{P} = \mathcal{P}_0 \otimes \mathcal{P}_0$. This last bound is sharp, in the sense that one cannot replace $p\sqrt{\log(1/p)}$ with a function of the order $o_{p \rightarrow 0}(p\sqrt{\log(1/p)})$ in the right-hand side, uniformly in G and ε .

Remark 91. The cut norm of a function on $X_1 \times X_2$ is defined as in (1), even if there is now a joint probability distribution on $X_1 \times X_2$ (which may not split as a product of distributions on X_1 and X_2); more precisely,

$$\|f\|_{\square} := \sup_{A_i \subset X_i} |\mathbf{E}[f \mathbb{1}_{A_1 \times A_2}]|.$$

Proof of Theorem 90. Let \mathcal{P} be the family of partitions of the form $\mathcal{P}_1 \otimes \mathcal{P}_2$, where \mathcal{P}_i is a partition of X_i . We apply Lemma 44 for the family \mathcal{P} , $\varepsilon^2/4$ in place of ε , the complexity function $\mathbb{C}(\mathcal{P}) = \log_2 |\mathcal{P}|$, and the energy function $\mathbf{E}_f^\nu(\mathcal{P})$; this produces a partition $\mathcal{P} = \mathcal{P}_1 \otimes \mathcal{P}_2$ with at most $2^{4/\varepsilon^2}$ parts, such that

$$\mathcal{E}_f^\nu(\mathcal{P} \wedge \mathcal{Q}) - \mathcal{E}_f^\nu(\mathcal{P}) \leq \frac{\varepsilon^2}{2} \mathbf{E} \left[f \log \frac{f}{\nu} \right] \log_2 |\mathcal{Q}|,$$

for any partition $\mathcal{Q} = \mathcal{Q}_1 \otimes \mathcal{Q}_2$ of $X_1 \times X_2$. Now consider the cut norm

$$\left\| f - \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \nu \right\|_{\square} = \sup_{A_i \subset X_i} \left| \mathbf{E} \left[f \mathbb{1}_{A_1 \times A_2} - \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \nu \mathbb{1}_{A_1 \times A_2} \right] \right|.$$

For fixed A_1 and A_2 , let $\mathcal{Q}_1 := \{A_1, A_1^c\}$, $\mathcal{Q}_2 := \{A_2, A_2^c\}$ and $\mathcal{Q} := \mathcal{Q}_1 \otimes \mathcal{Q}_2$. Then by Lemma 89, we have

$$\begin{aligned} & \left| \mathbf{E} \left[f \mathbb{1}_{A_1 \times A_2} - \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \nu \mathbb{1}_{A_1 \times A_2} \right] \right| \\ &= \left| \mathbf{E} \left[\mathbf{E}(f \mid \mathcal{P} \wedge \mathcal{Q}) \mathbb{1}_{A_1 \times A_2} - \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \mathbf{E}(\nu \mid \mathcal{P} \wedge \mathcal{Q}) \mathbb{1}_{A_1 \times A_2} \right] \right| \\ &\leq \mathbf{E} \left| \mathbf{E}(f \mid \mathcal{P} \wedge \mathcal{Q}) - \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \mathbf{E}(\nu \mid \mathcal{P} \wedge \mathcal{Q}) \right| \\ &\leq \sqrt{2 (\mathcal{E}_f^\nu(\mathcal{P} \wedge \mathcal{Q}) - \mathcal{E}_f^\nu(\mathcal{P}))} \\ &\leq \sqrt{2 \frac{\varepsilon^2}{4} \mathbf{E} \left[f \log \frac{f}{\nu} \right] \log_2 4}, \end{aligned}$$

which is what we wanted. When $X_1 = X_2$, one can instead use the family \mathcal{P} of partitions of the form $\mathcal{P}_0 \otimes \mathcal{P}_0$, and make the necessary modifications to \mathcal{Q} (i.e., take $\mathcal{Q}_0 := \mathcal{Q}_1 \wedge \mathcal{Q}_2$ and $\mathcal{Q} := \mathcal{Q}_0 \otimes \mathcal{Q}_0$).

The statement about graphs follows by letting $X_1 = X_2 := V(G)$, equipping $V(G) \times V(G)$ with the uniform probability measure, and letting $f := \mathbb{1}_G/p$ and $\nu \equiv 1$.

Finally, let us show that for any constant $C > 0$, one cannot in general choose a partition $\mathcal{P} = \mathcal{P}_0 \otimes \mathcal{P}_0$, where \mathcal{P}_0 contains at most $2^{C/\varepsilon^2}$ parts, such that

$$\|\mathbb{1}_G - \mathbf{E}(\mathbb{1}_G \mid \mathcal{P})\|_{\square} \leq \varepsilon o_{p \rightarrow 0} \left(p \sqrt{\log(1/p)} \right)$$

holds simultaneously for all $\varepsilon \in (0, 1)$ and all finite graphs G (where $p := \mathbf{E}[\mathbb{1}_G]$), allowing self-edges. Let $n \geq 1$ be large, $p := 1/n$, and $\varepsilon := \sqrt{C/\log_2(n/2)}$; we will construct a graph $G = G_n$ on n vertices, with edge density p , such that

$$\|\mathbb{1}_{G_n} - \mathbf{E}(\mathbb{1}_{G_n} \mid \mathcal{P})\|_{\square} \gg \frac{1}{n},$$

for all $\mathcal{P} = \mathcal{P}_0 \otimes \mathcal{P}_0$ such that \mathcal{P}_0 contains at most $2^{C/\varepsilon^2} = n/2$ sets. This will complete our proof since

$$\varepsilon p \sqrt{\log(1/p)} = \frac{1}{n} \sqrt{\frac{C \log n}{\log_2(n/2)}} \ll_C \frac{1}{n},$$

and thus the ratio $\|\mathbb{1}_G - \mathbf{E}(\mathbb{1}_G \mid \mathcal{P})\|_{\square} / (\varepsilon p \sqrt{\log(1/p)})$ does not vanish as $n \rightarrow \infty$ (i.e., $p \searrow 0$).

The construction is simple, using the same example as in [14, below Theorem 2.2]: take $V(G) = \{1, \dots, n\}$ and $\mathbb{1}_G(a, b) = \mathbb{1}_{a=b}$ for $a, b \in V(G)$, so that G consists of n loops. Then any partition \mathcal{P}_0 of $V(G)$ containing at most $n/2$ sets can use at most $n/2$ singletons (trivially); if $k_1, \dots, k_m \geq 2$ denote the sizes of the remaining sets A_1, \dots, A_m in \mathcal{P}_0 , we thus have

$$k_1 + \dots + k_m \geq n - \frac{n}{2} = \frac{n}{2}.$$

Now consider a cut $C \subset V(G)$ which splits each A_i roughly in half; more precisely, choose $C \subset A_1 \cup \dots \cup A_m$ such that $|C \cap A_i| = \lceil k_i/2 \rceil$ for $i \in \{1, \dots, m\}$. Then we have

$$\begin{aligned} \|\mathbb{1}_{G_n} - \mathbf{E}(\mathbb{1}_{G_n} \mid \mathcal{P})\|_{\square} &\geq \mathbf{E}[(\mathbb{1}_{G_n} - \mathbf{E}(\mathbb{1}_{G_n} \mid \mathcal{P})) \mathbb{1}_{C \times C}] \\ &= \frac{1}{n^2} \sum_{i=1}^m \sum_{a,b \in C \cap A_i} \left(\mathbb{1}_{a=b} - \frac{k_i}{k_i^2} \right) \\ &= \frac{1}{n^2} \sum_{i=1}^m \left(|C \cap A_i| - \frac{|C \cap A_i|^2}{k_i} \right) = \frac{1}{n^2} \sum_{i=1}^m \left\lceil \frac{k_i}{2} \right\rceil \left(1 - \frac{\lceil k_i/2 \rceil}{k_i} \right). \end{aligned}$$

Since each $k_i \geq 2$, it is easy to check that $\lceil k_i/2 \rceil/k_i \leq 2/3$, and thus

$$\|\mathbb{1}_{G_n} - \mathbf{E}(\mathbb{1}_{G_n} \mid \mathcal{P})\|_{\square} \gg \frac{1}{n^2} \sum_{i=1}^m \frac{k_i}{2} \gg \frac{1}{n},$$

which completes our proof. \square

Remark 92. One can also prove a *strong* relative regularity lemma similar to Theorem 90 using an additional level of iteration; we leave the details to the interested reader.

The corollary below should be compared to the dense model theorem from [17], noting the dependency of ε' on ε and δ .

Corollary 93 (Dense model theorem). *For any $\varepsilon \in (0, 1)$, there exists $\varepsilon' > 0$ such that the following holds. Let $(X_1 \times X_2, \mathbf{P})$ be a finite probability space and $f, \nu : X_1 \times X_2 \rightarrow [0, \infty)$ satisfy*

$$f \leq \nu \quad \text{and} \quad \|\nu - 1\|_{\square} \leq \varepsilon'.$$

Then there exists a function $\tilde{f} : X_1 \times X_2 \rightarrow [0, \infty)$ such that

$$\tilde{f} \leq 1 \quad \text{and} \quad \|f - \tilde{f}\|_{\square} \leq \varepsilon.$$

In fact, if $\delta := \mathbf{E}[f] > 0$ and $\mathbf{E}[\nu] = 1$, then one can take $\varepsilon' \gg \varepsilon \exp(-O(\delta^2 \log \frac{1}{\delta}) \varepsilon^{-2})$.

Proof. If $f \equiv 0$, there is nothing to prove; taking $\varepsilon' < 1/2$, we can also guarantee that $\mathbf{E}[\nu] \asymp 1$. So let $\delta := \mathbf{E}[f] > 0$, $g := f/\delta$, and $\mu := \nu/\mathbf{E}[\nu]$. Then by Theorem 90, for any $\tau \in (0, 1)$, there exists a partition $\mathcal{P} = \mathcal{P}_1 \otimes \mathcal{P}_2$ of $X_1 \times X_2$ with $2^{O(1/\tau^2)}$ parts such that

$$\left\| g - \frac{\mathbf{E}(g \mid \mathcal{P})}{\mathbf{E}(\mu \mid \mathcal{P})} \mu \right\|_{\square} \leq \tau \sqrt{\mathbf{E} \left[g \log \frac{g}{\mu} \right]}.$$

Renormalizing and using the inequality $f \leq \mu \mathbf{E}[\nu]$, this translates to

$$\left\| f - \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} \nu \right\|_{\square} \leq \delta \tau \sqrt{\mathbf{E} \left[\frac{f}{\delta} \log \frac{f}{\delta \mu} \right]} \leq \delta \tau \sqrt{\log \frac{\mathbf{E}[\nu]}{\delta}}.$$

On the other hand, since $\mathbf{E}(f \mid \mathcal{P}) \leq \mathbf{E}(\nu \mid \mathcal{P})$, one can obtain the inequality

$$\left\| \frac{\mathbf{E}(f \mid \mathcal{P})}{\mathbf{E}(\nu \mid \mathcal{P})} (\nu - 1) \right\|_{\square} \leq 2^{O(1/\tau^2)} \|\nu - 1\|_{\square},$$

by expanding the right-hand side as a sum of $2^{O(1/\tau^2)}$ terms (corresponding to the parts of $\mathcal{P}_1 \otimes \mathcal{P}_2$).

Putting the last two inequalities together and assuming that $\|\nu - 1\|_{\square} \leq \varepsilon'$ (for some $\varepsilon' > 0$ to be chosen), we get that

$$\|f - \tilde{f}\|_{\square} \leq 2^{O(1/\tau^2)} \varepsilon' + \delta \tau \sqrt{\log \frac{\mathbf{E}[\nu]}{\delta}},$$

where $\tilde{f} := \mathbf{E}(f \mid \mathcal{P})/\mathbf{E}(\nu \mid \mathcal{P}) \leq 1$. The conclusion follows by choosing τ and then ε' small enough such that each of the above terms is at most $\varepsilon/2$. \square

Acknowledgements

The author wishes to thank his PhD advisor, Professor James Maynard, for his guidance and support, and Professors Ben Green, Terence Tao, Yufei Zhao, Benny Sudakov, and Peter Allen, for helpful discussions. This work was completed while the author was sponsored by the EPSRC Scholarship at the University of Oxford.

References

- [1] Peter Allen, Julia Böttcher, Jozef Skokan, and Maya Stein. Regularity inheritance in pseudorandom graphs. *Random Structures Algorithms*, 56(2):306–338, 2020.
- [2] Peter Allen, Peter Keevash, Benny Sudakov, and Jacques Verstraëte. Turán numbers of bipartite graphs plus an odd cycle. *J. Combin. Theory Ser. B*, 106:134–162, 2014.
- [3] Noga Alon, Lajos Rónyai, and Tibor Szabó. Norm-graphs: variations and applications. *J. Combin. Theory Ser. B*, 76(2):280–290, 1999.
- [4] Keith Ball, Eric A. Carlen, and Elliott H. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Invent. Math.*, 115(3):463–482, 1994.
- [5] József Balogh, Robert Morris, and Wojciech Samotij. Independent sets in hypergraphs. *J. Amer. Math. Soc.*, 28(3):669–709, 2015.
- [6] Rajendra Bhatia and Fuad Kittaneh. Cartesian decompositions and Schatten norms. *Linear Algebra Appl.*, 318(1-3):109–116, 2000.
- [7] George R. Blakley and Prabir Roy. A Hölder type inequality for symmetric matrices with nonnegative entries. *Proc. Amer. Math. Soc.*, 16:1244–1245, 1965.
- [8] Greg Bodwin and Santosh Vempala. A unified view of graph regularity via matrix decompositions. *Random Structures Algorithms*, 61(1):62–83, 2022.
- [9] Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao. An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.*, 46(1):337–396, 2018.
- [10] Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao. An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.*, 372(5):3019–3062, 2019.
- [11] Fan R. K. Chung, Ronald L. Graham, and Richard M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.
- [12] Sebastian M. Cioabă. Closed walks and eigenvalues of abelian Cayley graphs. *C. R. Math. Acad. Sci. Paris*, 342(9):635–638, 2006.
- [13] David Conlon and Jacob Fox. Graph removal lemmas. In *Surveys in combinatorics 2013*, volume 409 of *London Math. Soc. Lecture Note Ser.*, pages 1–49. Cambridge Univ. Press, Cambridge, 2013.
- [14] David Conlon, Jacob Fox, Benny Sudakov, and Yufei Zhao. The regularity method for graphs with few 4-cycles. *J. Lond. Math. Soc. (2)*, 104(5):2376–2401, 2021.

- [15] David Conlon, Jacob Fox, Benny Sudakov, and Yufei Zhao. Which graphs can be counted in C_4 -free graphs? *Pure Appl. Math. Q.*, 18(6):2413–2432, 2022.
- [16] David Conlon, Jacob Fox, and Yufei Zhao. Extremal results in sparse pseudorandom graphs. *Adv. Math.*, 256:206–290, 2014.
- [17] David Conlon, Jacob Fox, and Yufei Zhao. The Green-Tao theorem: an exposition. *EMS Surv. Math. Sci.*, 1(2):249–282, 2014.
- [18] David Conlon, William T. Gowers, Wojciech Samotij, and Mathias Schacht. On the KŁR conjecture in random graphs. *Israel J. Math.*, 203(1):535–580, 2014.
- [19] David Conlon and Joonkyung Lee. Finite reflection groups and graph norms. *Adv. Math.*, 315:130–165, 2017.
- [20] Keith Dsouza and Mike Krebs. A combinatorial trace method: counting closed walks to assay graph eigenvalues. *Rocky Mountain J. Math.*, 43(2):469–478, 2013.
- [21] Paul Erdős, Peter Frankl, and Vojtech Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.*, 2(2):113–121, 1986.
- [22] Jacob Fox. A new proof of the graph removal lemma. *Ann. of Math. (2)*, 174(1):561–579, 2011.
- [23] Jacob Fox, László Miklós Lovász, and Yufei Zhao. A fast new algorithm for weak graph regularity. *Combin. Probab. Comput.*, 28(5):777–790, 2019.
- [24] Alan Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.
- [25] Alan Frieze and Ravi Kannan. A simple algorithm for constructing Szemerédi’s regularity partition. *Electron. J. Combin.*, 6:#17, 1999.
- [26] Zoltán Füredi. An upper bound on Zarankiewicz’ problem. *Combin. Probab. Comput.*, 5(1):29–33, 1996.
- [27] William T. Gowers. A new proof of Szemerédi’s theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.
- [28] William T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. of Math. (2)*, 166(3):897–946, 2007.
- [29] Ben Green. Roth’s theorem in the primes. *Ann. of Math. (2)*, 161(3):1609–1636, 2005.
- [30] Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008.
- [31] Hamed Hatami. *On generalizations of Gowers norms*. ProQuest LLC, Ann Arbor, MI, 2009. Thesis (Ph.D.)—University of Toronto (Canada).
- [32] Tao Jiang, Sean Longbrake, and Jie Ma. Bipartite-ness under smooth conditions. *Combin. Probab. Comput.*, 32(4):546–558, 2023.

- [33] Yoshiharu Kohayakawa. Szemerédi's regularity lemma for sparse graphs. In *Foundations of computational mathematics (Rio de Janeiro, 1997)*, pages 216–230. Springer, Berlin, 1997.
- [34] Yoshiharu Kohayakawa, Tomasz Łuczak, and Vojtech Rödl. On K^4 -free subgraphs of random graphs. *Combinatorica*, 17(2):173–213, 1997.
- [35] László Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.
- [36] László Lovász and Balázs Szegedy. Szemerédi's lemma for the analyst. *Geom. Funct. Anal.*, 17(1):252–270, 2007.
- [37] Vladimir Nikiforov. Extremal norms of graphs and matrices. *J. Math. Sci. (N.Y.)*, 182(2):164–174, 2012. Translated from Sovrem. Mat. Prilozh., Vol. 71, 2011.
- [38] Sean Prendiville. Solving equations in dense Sidon sets. *Math. Proc. Cambridge Philos. Soc.*, 173(1):25–34, 2022.
- [39] Klaus F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.
- [40] Imre Z. Ruzsa and Endre Szemerédi. Triple systems with no six points carrying three triangles. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, volume 18 of *Colloq. Math. Soc. János Bolyai*, pages 939–945. North-Holland, Amsterdam-New York, 1978.
- [41] David Saxton and Andrew Thomason. Hypergraph containers. *Invent. Math.*, 201(3):925–992, 2015.
- [42] Alexander Scott. Szemerédi's regularity lemma for matrices and sparse graphs. *Combin. Probab. Comput.*, 20(3):455–466, 2011.
- [43] Alexander F. Sidorenko. Inequalities for functionals generated by bipartite graphs. *Diskret. Mat.*, 3(3):50–65, 1991.
- [44] József Solymosi. Note on a generalization of Roth's theorem. In *Discrete and computational geometry*, volume 25 of *Algorithms Combin.*, pages 825–827. Springer, Berlin, 2003.
- [45] József Solymosi. A note on a question of Erdős and Graham. *Combin. Probab. Comput.*, 13(2):263–267, 2004.
- [46] Balázs Szegedy. Limits of kernel operators and the spectral regularity lemma. *European J. Combin.*, 32(7):1156–1167, 2011.
- [47] Endre Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.
- [48] Terence Tao. The Gaussian primes contain arbitrarily shaped constellations. *J. Anal. Math.*, 99:109–176, 2006.
- [49] Terence Tao. Szemerédi's regularity lemma revisited. *Contrib. Discrete Math.*, 1(1):8–28, 2006.

- [50] Terence Tao. A variant of the hypergraph removal lemma. *J. Combin. Theory Ser. A*, 113(7):1257–1280, 2006.
- [51] Terence Tao. The ergodic and combinatorial approaches to Szemerédi’s theorem. In *Additive combinatorics*, volume 43 of *CRM Proc. Lecture Notes*, pages 145–193. Amer. Math. Soc., Providence, RI, 2007.
- [52] Terence Tao and Tamar Ziegler. A multi-dimensional Szemerédi theorem for the primes via a correspondence principle. *Israel J. Math.*, 207(1):203–228, 2015.
- [53] Xingzhi Zhan. *Matrix inequalities*, volume 1790 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002.
- [54] Antoni Zygmund. *Trigonometric series. Vol. I, II*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2002. With a foreword by Robert A. Fefferman.