Efficient Arithmetic Regularity and Removal Lemmas for Induced Bipartite Patterns

Noga Alon[*]   Jacob Fox[†]   Yufei Zhao[‡]

Abstract: Let G be an abelian group of bounded exponent and $A \subseteq G$. We show that if the collection of translates of A has VC dimension at most d, then for every $\varepsilon > 0$ there is a subgroup H of G of index at most $\varepsilon^{-d-o(1)}$ such that one can add or delete at most $\varepsilon|G|$ elements to/from A to make it a union of H-cosets.

We also establish a removal lemma with polynomial bounds, with applications to property testing, for induced bipartite patterns in a finite abelian group with bounded exponent.

Key words and phrases: regularity lemma, removal lemma, property testing, VC dimension, arithmetic regularity, induced patterns

1   Introduction

Szemerédi's regularity lemma [31] gives a rough structural decomposition for all graphs and is one of the most powerful tools in graph theory. A major drawback of the regularity lemma is that the number of parts in the decomposition grows as an exponential tower of 2's of height a power of $1/\varepsilon$, where ε is the regularity parameter [12]. A natural question that has been studied by many researchers is: in what circumstances can one get a more effective bound? Namely, under what conditions does every graph in a family $\mathcal{F}$ of graphs necessarily have a partition with much fewer parts, say polynomial in $1/\varepsilon$? One natural condition for a family $\mathcal{F}$ of graphs is that it is *hereditary*, that is, if $G \in \mathcal{F}$ then every induced subgraph of G is also in $\mathcal{F}$. For hereditary families, it turns out that the bound on the number of parts in a regular partition is polynomial in $1/\varepsilon$ if the neighborhood set system of every graph in the family has

bounded VC dimension, and otherwise the bound is tower-type. This gives a satisfactory answer to the problem.

A *set system* $\mathcal{S}$ is a collection of subsets of some ground set Ω. Here we only consider finite Ω. We say that $U \subseteq \Omega$ is *shattered* by $\mathcal{S}$ if for every $U' \subseteq U$ there is some $T \in \mathcal{S}$ with $T \cap U = U'$. The *Vapnik–Chervonenkis dimension* (or *VC dimension*) of $\mathcal{S}$, denoted VC-dim$\mathcal{S}$, is the size of the largest shattered set.

Let G be a graph. The *neighborhood $N(v)$* of a vertex $v \in V(G)$ is the set of vertices adjacent to v. The *VC dimension* of a graph G is defined to be VC-dim$\{N(v) : v \in V\}$.

Given a bipartite graph F with vertex bipartition $V(F) = U \cup V$, we say that a map $\phi \colon V(F) \to V(G)$ *bi-induces F* if for every $(u, v) \in U \times V$, the pair uv is an edge of F if and only if $\phi(u)\phi(v)$ is an edge of G. Note that we have no requirements about edges in G between vertices in the image of U, and likewise with V. We say that G contains a *bi-induced copy of H* if there exists a map ϕ as above that is injective on each of U and V.[1]

It is known that the following are equivalent for a hereditary family $\mathcal{F}$ of graphs:

(1) The VC dimension of the graphs in $\mathcal{F}$ is uniformly bounded.

(2) There is a bipartite graph F such that none of the graphs in $\mathcal{F}$ has a bi-induced copy of F.

(3) The family $\mathcal{F}$ has a forbidden induced bipartite graph, a forbidden induced complement of a bipartite graph, and a forbidden induced split graph.

(4) The number of graphs in $\mathcal{F}$ on n vertices is at most $2^{n^{2-\varepsilon}}$ for some $\varepsilon = \varepsilon(\mathcal{F}) > 0$. In contrast, every other hereditary family of graphs contains at least $2^{n^2/4}$ labeled graphs on n vertices.

(5) There is a constant $k = k(\mathcal{F})$ such that every n-vertex graph in $\mathcal{F}$ has an equitable vertex partition into at most ε^{-k} parts such that all but at most an ε-fraction of the pairs of parts have edge density at most ε or at least $1 - \varepsilon$. In contrast, every other hereditary family of graphs has a graph that requires a tower in a power of $1/\varepsilon$ parts in any ε-regular equitable vertex partition.

The above characterizations give an interesting dichotomy between hereditary families of graphs of bounded VC dimension versus those of unbounded VC dimension. It shows that families of graphs with bounded VC dimension have smaller growth and are more structured. The equivalence of (1) and (4) was given by Alon, Balogh, Bollobás, and Morris [1]. Alon, Fischer, and Newman [3] proved a bipartite version of the regularity lemma for graphs of bounded VC dimension, and the version for all graphs is due to Lovász and Szegedy [19]. The proof was simplified with improved bounds by Fox, Pach, and Suk [11]. Further results related to the above equivalences for tournaments can be found in [10].

A *half-graph* is a bipartite graph on $2k$ vertices $\{u_1, \ldots, u_k\} \cup \{v_1, \ldots, v_k\}$ such that u_i is adjacent to v_j if and only if $i \leq j$. Malliaris and Shelah [20] proved if a graph has no bi-induced copy of the half-graph on $2k$ vertices, then one can partition the vertex set into $\varepsilon^{-O_k(1)}$ many parts such that every pair of parts

---

[1]Having a bi-induced copy of F is weaker than having an *induced* copy of F, where in the latter we also require that there are no edges in G between vertices in the image of U, and likewise with V. Also, an alternative notion of bi-induced copy of H assumes that ϕ is injective. The discussed results hold for this alternative notion as well.

is ε-regular (there are no irregular pairs). Bi-inducing a half-graph is related to a notion of stability in model theory, and for this reason Malliaris and Shelah called their result a "stable regularity lemma".

The above discussion summarizes some relevant results for graphs. We now turn our attention to subsets of groups and their associated Cayley graphs. Let G be a finite abelian group, written additively. Let $A \subseteq G$. Consider the *Cayley sum graph* formed by taking the elements of G as vertices, where $x, y \in G$ are adjacent if $x + y \in A$ (we may end up with some loops; alternatively, we can consider a bipartite version of this construction). The VC dimension of the graph corresponds to the VC dimension of the collection of translates of A, which we simply call the *VC dimension of A*, defined as

$$\text{VC-dim}A := \text{VC-dim}\{A + x : x \in G\}.$$

For a bipartite graph F with vertex bipartition $U \cup V$, we say that a map $\phi: V(F) \to G$ *bi-induces F in A* if, for every $(u, v) \in U \times V$, uv is an edge of F if and only if $\phi(u) + \phi(v) \in A$. We say that A has a *bi-induced copy of F* if there exists a map ϕ as above that is injective on each of U and V.

Observe that A has a bi-induced copy of F if its VC dimension is large enough. To see this, first note that if no pair of vertices in U have identical neighborhoods in V, and A has VC dimension at least $|V|$, then A has a bi-induced copy of F. Indeed, we can construct ϕ by mapping V to a subset of G shattered by translates of A (such a choice exists since VC-dim$A \geq |V|$). Since $\phi(V)$ is shattered, for every $u \in U$, there is some $y_u \in G$ such that $(A - y) \cap \phi(V) = \phi(N(u))$. Let ϕ send u to this y_u, for each $u \in U$. We obtain a map $\phi: V(F) \to G$ that bi-induces F, though this map may not be injective on U (it is always injective on V) if some pairs of vertices of U have identical neighborhoods, but this can be easily fixed[2].

Green [13] proved an arithmetic analogue of Szemerédi's regularity lemma for abelian groups. The statement is much simpler in the case of abelian groups of bounded exponents, which is the main focus of our paper (some remarks regarding general groups are given in the final section). For an abelian group G and a subset $A \subseteq G$, a coset $H + x$ of a subgroup H is called ε-regular if all the nontrivial Fourier coefficients of $A \cap (H + x)$, when interpreted as a subset of $H + x$, are at most ε. For each $\varepsilon > 0$ and positive integer r, Green's arithmetic regularity lemma states that there is $K = K(r, \varepsilon)$ such that the following holds. If G has exponent at most r and $A \subseteq G$, then there is a subgroup $H \subseteq G$ of index at most K such that all but an ε-fraction of the cosets of H are ε-regular.

Recently, an arithmetic analog of the Malliaris–Shelah stable regularity lemma was proved by Terry and Wolf [30] for $G = \mathbb{F}_p^n$ with p fixed. It was shown that if $A \subseteq G$ has no bi-induced copy of a half-graph on $2k$ vertices, then there is a subgroup H of G of index at most $e^{\varepsilon^{-O_{k,p}(1)}}$ such that for every $x \in G$, one has either $|A \cap (H + x)| \leq \varepsilon |H|$ or $|A \cap (H + x)| \geq (1 - \varepsilon) |H|$. Here the subscripts on the $O_{k,p}(1)$ mean that the constant is allowed to depend on k and p. The result was subsequently extended to general groups by Conant, Pillay, and Terry [6], who showed that for every finite group G, if $A \subseteq G$ has no bi-induced copy of the half-graph on $2k$ vertices, then there is a normal subgroup H of G of index $O_{k,\varepsilon}(1)$ such that there is some union S of H-cosets such that $|A \Delta S| \leq \varepsilon |H|$, where $A \Delta B = (A \setminus B) \cup (B \setminus A)$ denotes the symmetric difference. However, the general group version of the theorem [6] gives no quantitative bounds on the index of H due to the model theoretic tools involved in its proof.

---

[2]Consider the bipartite graph F_+ obtained from F by adding $\lceil \log_2 |U| \rceil$ new vertices to the vertex set V, and add edges from the new vertices to U so that no two vertices in U have identical neighborhoods in F_+. By earlier arguments, if VC-dim$A \geq |V| + \lceil \log_2 |V| \rceil$, then A necessarily contains an bi-induced copy of F_+, and hence a bi-induced copy of F.

We saw earlier that forbidding a fixed bi-induced bipartite graph implies bounded VC dimension. Our first main result generalizes a variant of Terry and Wolf's result to sets of bounded VC dimension, and gives bounds of polynomial order in $1/\varepsilon$. Its proof can be found in Section 2.

Theorem 1.1 (Regularity lemma). *Fix positive integers r and d. If G is a finite abelian group with exponent at most r, and $A \subseteq G$ has VC dimension at most d, then for every $\varepsilon > 0$ there is a subgroup H of G of index at most $\varepsilon^{-d-o(1)}$ such that $|A \Delta S| \leq \varepsilon |G|$ for some $S \subseteq G$ which is a union of cosets of H.*

Here $o(1)$ is some quantity that goes to zero as $\varepsilon \to 0$, at a rate possibly depending on r and d.

We also prove a removal lemma for bi-induced copies of a fixed bipartite graph. Let us first recall the classical graph removal lemma. We say that an n-vertex graph is ε-*far* from some property if one needs to add or delete more than εn^2 edges to satisfy the property. The triangle removal lemma[3] says that if an n-vertex graph is ε-far from triangle-free, then its triangle density is at least $\delta(\varepsilon) > 0$. The original graph regularity proof [23] of the triangle removal lemma shows that we may take $1/\delta(\varepsilon)$ to be a tower of 2's of height $\varepsilon^{-O(1)}$, which was improved to height $O(\log(1/\varepsilon))$ in [9]. It is known that there exists a constant $c > 0$ such that the bound in the triangle removal lemma cannot be improved to $\delta = \varepsilon^{-c\log(1/\varepsilon)}$ (see [8] for a survey on graph removal lemmas). There is also a removal lemma for induced subgraphs [2], initially proved using a so-called *strong regularity lemma*, though better bounds were later obtained in [7].

An arithmetic analog of the graph removal lemma was first proved by Green [13] for "complexity 1" patterns such as $x + y + z = 0$ using his arithmetic regularity lemma. Král', Serra, and Vena [17] later showed that Green's arithmetic removal lemma can be deduced as a consequence of the graph removal lemma. More general arithmetic removal lemmas for linear systems were later proved as a consequence of the hypergraph removal lemma [18, 27]. We refer to the references for precise statements. Note that the reduction from the arithmetic removal lemma to the (hyper)graph removal lemma fails for induced patterns. It remains open to find a general induced arithmetic removal lemmas [27, Conjecture 5.3].

Our second main result gives an arithmetic analog of the removal lemma, with polynomial bounds, for bi-induced patterns. We say that $A \subset G$ is ε-*far from bi-induced-F-free* if $A' \subset G$ contains a bi-induced copy of F whenever $|A \Delta A'| \leq \varepsilon |G|$. Here is our second main result, whose proof can be found in Section 4.

Theorem 1.2 (Removal lemma). *Fix a positive integer r and a bipartite graph F. Let G be a finite abelian group with exponent at most r. For every $0 < \varepsilon < 1/2$, if $A \subseteq G$ is ε-far from bi-induced-F-free, then the probability that a uniform random map $\phi \colon V(F) \to G$ bi-induces F is at least $\varepsilon^{O(|V(F)|^3)}$.*

We mention an application to property testing. The removal lemma gives a polynomial-time randomized sampling algorithm for distinguishing sets $A \subseteq G$ that are bi-induced-F-free from those that are ε-far from bi-induced-F-free. Indeed, sample a random map $\phi \colon V(F) \to G$, and output YES if ϕ bi-induces F and is injective on each vertex part of F, and otherwise output NO. If A is bi-induced-F-free, then the algorithm always outputs NO. On the other hand, if F is ε-far from bi-induced-F-free, then by the theorem above, the algorithm outputs YES with probability at least $\varepsilon^{O_F(1)}$, provided that G is large enough, so that ϕ is injective with high probability. We can then repeat the experiment $\varepsilon^{-O_F(1)}$ times to obtain a randomized algorithm that succeeds with high probability.

---

[3]The removal lemma is often stated in the contrapositive, which better explains the name "removal lemma": if triangle density of a graph is at most $\delta(\varepsilon) > 0$, then the graph can be made triangle-free by deleting εn^2 edges

2 Regularity lemma

In this section, we prove Theorem 1.1.

We say that a set system $\mathcal{S}$ on a finite ground set Ω is δ-*separated* if $|S \Delta T| \geq \delta |\Omega|$ for all distinct $S, T \in \mathcal{S}$. We quote a bound on the size of a δ-separated system.

Lemma 2.1 (Haussler's packing lemma [15])**.** *Let $d, \delta > 0$. If $\mathcal{S}$ is a δ-separated set system of VC dimension at most d, then $|\mathcal{S}| \leq (30/\delta)^d$.*

By taking a maximal δ-separated collection of translates of $A \subseteq G$, we deduce, below, that A must be δ-close to many of its own translates.

Lemma 2.2. *Let G be a finite abelian group, and $A \subseteq G$ a subset with VC dimension at most d, and $\delta > 0$. Then*

$$|\{x : |A \Delta (A + x)| \leq \delta |G|\}| \geq (\delta/30)^d |G|.$$

Proof. Let W be a maximal subset of G such that $|(A + w) \Delta (A + w')| > \delta |G|$ for all distinct $w, w' \in W$. We have $|W| \leq (30/\delta)^d$ by Lemma 2.1. Let

$$B = \{x \in G : |A \Delta (A + x)| \leq \delta |G|\}.$$

Since W is maximal, for every $x \in G$, there is some $w \in W$ such that $|(A + x) \Delta (A + w)| \leq \delta |G|$, which implies $x - w \in B$. Hence $G = \bigcup_{w \in W} (B + w)$. Therefore $|B| \geq |G| / |W| \geq (\delta/30)^d |G|$. $\qquad \square$

We quote a result from additive combinatorics. We use the following standard notation: $A + A = \{a + b : a, b \in A\}$, $A - A = \{a - b : a, b \in A\}$, and $kA = A + \cdots + A$ (k times).

Theorem 2.3 (Bogolyubov–Ruzsa lemma for groups with bounded exponent)**.** *Let G be an abelian group of exponent at most r, and $A \subseteq G$ a finite subset with $|A + A| \leq K |A|$. Then $2A - 2A$ contains a subgroup of G of size at least $c_r(K) |A|$ for some constant $c_r(K) > 0$.*

The name "Bogolyubov–Ruzsa lemma" was given by Sanders [24], who proved the theorem with the current best bound $c_r(K) = e^{-O_r(\log^4 2K)}$ (see [24, Theorem 11.1]). We refer the readers to the introductions of [24, 25] for the history of this result. A version of the theorem for $G = \mathbb{Z}$ was initially proved by Ruzsa [21] as a key step towards his proof of Freiman's theorem. The assertion of the polynomial Freiman–Ruzsa conjecture, a central open problem in additive combinatorics, would follow from an improvement of the bound to $c_r(K) = K^{-O_r(1)}$.

In our next lemma, we start from the conclusion of Lemma 2.2, which gives us a large set B such that $A \approx A + x$ for all $x \in B$. Consider the sequence $B, 2B, 4B, 8B, \ldots$. Since B is large, the size of $2^i B$ cannot keep on growing, so we can find a set $B' = 2^i B$ with small doubling $|B' + B'| \leq K |B'|$, and i not too large. Theorem 2.3 then implies that $2B' - 2B'$ contains a large subgroup, in which every element x satisfies $A \approx A + x$, which is close to what we need.

Lemma 2.4. *Fix a positive integer r. Let G be a finite abelian group of exponent at most r. Let $0 < \delta < 1/2$, $C > 0$, and $A \subseteq G$. Let $B = \{x \in G : |A \Delta (A + x)| \leq \delta |G|\}$. Suppose $|B| \geq \delta^C |G|$. Then there exists a subgroup H of G with $|H| \geq \delta^{o(1)} |B|$ such that $|A \Delta (A + x)| \leq \delta^{1 - o(1)} |G|$ for all $x \in H$, and furthermore there exists a union S of H-cosets such that $|A \Delta S| \leq \delta^{1 - o(1)} |G|$. Here $o(1)$ is a quantity that goes to zero as $\delta \to 0$, at a rate that may depend on r and C.*

Proof. Let $K = K(\delta) > 1$ to be decided. We cannot have $|2^{i+1}B| > K|2^iB|$ for every $0 \le i \le \log_K(|G|/|B|)$ since otherwise we would have $|2^iB| > |G|$ for some i, which is impossible as 2^iB is a subset of G. Thus $|2^{i+1}B| \le K|2^iB|$ for some $i \le \log_K(|G|/|B|) \le C\log_K(1/\delta)$, and letting $\ell = 2^i$, we have

$$|2\ell B| \le K|\ell B| \quad \text{with} \quad \ell \le 2^{C\log(1/\delta)/\log K} = \delta^{-O(1/\log K)}. \tag{1}$$

Since $|(A+x)\Delta A| \le \delta|G|$ for all $x \in B$, we have, by the triangle inequality,

$$|(A+x+y)\Delta A| \le |(A+x+y)\Delta(A+y)| + |(A+y)\Delta A| = |(A+x)\Delta A| + |(A+y)\Delta A| \text{ for all } x,y \in G.$$

Thus

$$|A\Delta(A+x)| \le 4\ell\delta|G| \quad \text{for all } x \in 2\ell B - 2\ell B. \tag{2}$$

By Theorem 2.3 and (1), $2\ell B - 2\ell B$ contains a subgroup H of G with $|H| \ge c_r(K)|\ell B| \ge c_r(K)|B|$. This would complete the proof of the first claim in the lemma provided that $K = K(\delta) \to \infty$ slowly enough as $\delta \to 0$ so that $c_r(K) = \delta^{o(1)}$ (then $\ell \le \delta^{-O(1/\log K)} = \delta^{-o(1)}$). Concretely, Theorem 2.3 with Sander's $c_r(K) = e^{-O_r(\log^4 2K)}$ allows us to take $K(\delta) = \exp((\log 1/\delta)^{1/5})$, say, so that all the $o(1)$'s in the exponents decay as $(\log(1/\delta))^{-1/5}$.

For the second claim, let S be the union of all H-cosets $y+H$ with $|A \cap (y+H)| \ge |H|/2$. Then

$$\begin{aligned}|A\Delta S| &= \sum_{y \in G/H} \min\{|A\cap(y+H)|, |H| - |A\cap(x+H)|\} \\ &\le \sum_{y \in G/H} \frac{2}{|H|}|A\cap(y+H)|(|H| - |A\cap(y+H)|) \\ &= \frac{1}{|H|}\sum_{x\in H}|A\Delta(A+x)| \quad \text{[counting pairs in } A\times(G\setminus A) \text{ lying in the same } H\text{-coset]} \\ &\le 4\ell\delta|G| = \delta^{1-o(1)}|G|. \quad \text{[by (2)]}\end{aligned}$$

$\square$

The regularity lemma, Theorem 1.1, then follows immediately after combining Lemmas 2.2 and 2.4.

Instead of applying the Bogolyubov–Ruzsa lemma as we do above, it is also possible to prove Lemma 2.4 using Freiman's theorem for groups of bounded exponent:

Theorem 2.5 (Ruzsa [22]). *If A is a finite subset of an abelian group of exponent at most r such that $|A+A| \le K|A|$, then A is contained in a subgroup of size $O_{r,K}(1)|A|$.*

At the point in the proof of Lemma 2.4 where we apply Theorem 2.3, we can instead apply Theorem 2.5 to contain ℓB inside a subgroup of size $\delta^{-o(1)}|\ell B|$. Now we apply a corollary of Kneser's theorem.

Theorem 2.6 (Kneser's theorem [16]; see [29, Theorem 5.5]). *Let G be an abelian group and A, B finite non-empty subsets. If $|A| + |B| \le |G|$ then there is a finite subgroup H of G such that*

$$|A+B| \ge |A+H| + |B+H| - |H| \ge |A| + |B| - |H|.$$

The subgroup H can be taken to be the stabilizer of $A+B$:

$$H = \{g \in G : g + (A+B) = (A+B)\}.$$

Corollary 2.7. *If G is an abelian group, t is a positive integer, and $A \subset G$ has $|A| \geq |G|/t$ and A generates G, then $2tA = G$.*

Proof. For any i such that $(i+1)A \neq G$, applying Kneser's theorem to the sets iA and A gives us a subgroup H so that $|(i+1)A| \geq |iA+H| + |A+H| - |H| \geq |iA| + |A|/2$ (since A generates G, $A+H$ is a union of at least two cosets of H, so $|H| \leq |A+H|/2$ and $|A+H| \geq |A|$). Iterating gives $|2tA| \geq t|A| \geq |G|$. □

Let us continue with our discussion of the alternative approach to proving Lemma 2.4. Since ℓB occupies a $\delta^{o(1)}$-fraction of some subgroup, by the above corollary, $\ell'B$ is a subgroup (playing the role of H in the first proof) for some $\ell' = \delta^{-o(1)}\ell$. From this point we can proceed as the rest of the proof of Lemma 2.4.

3 A strengthened regularity lemma

In the next section, we prove a removal lemma for bi-induced patterns. The regularity lemma we stated in Theorem 1.1 seems not quite strong enough to establish the removal lemma. Below we prove a strengthening, where the VC dimension hypothesis is weakened to a more robust one. Instead of requiring that A has bounded VC dimension, we will ask that, with probability at least 0.9, say, the VC dimension of the collection of translates of A is bounded if we restrict the ground set G to a random set. We state the result below in the form of two alternatives: either A has high VC dimension when sampled, or it satisfies a regularity lemma with polynomial bounds.

Proposition 3.1 (Regularity lemma with robust VC dimension hypothesis)**.** *Fix positive integers r and d. Let G be a finite abelian group of exponent at most r. Let $A \subseteq G$. One of the following must be true for every small $\varepsilon > 0$:*

(a) *For some $k = \varepsilon^{-d-o(1)}$, if X and Y are random k-element subset of G, then we have $\text{VC-dim}\{(A + x) \cap Y : x \in X\} > d$ with probability at least 0.9.*

(b) *There exists a subgroup H of G of index at most $\varepsilon^{-d-o(1)}$ such that $|A \Delta S| \leq \varepsilon |G|$ for some union S of H-cosets.*

Here $o(1)$ refers to a quantity that goes to zero as $\varepsilon \to 0$, at a rate that can depend on r and d.

Recall that Lemma 2.2 tells us that if $\text{VC-dim} A \leq d$, then $B = \{x : |A \Delta (A + x)| \leq \delta |G|\}$ has size at least $(\delta/30)^d |G|$. We will derive a similar bound for B under the weaker hypothesis, namely the negation of (a), from which we can deduce (b) using Lemma 2.4 as in the proof of the previous regularity lemma Theorem 1.1.

Lemma 3.2. *Let $k \leq n/2$ be positive integers. In an n-vertex graph with maximum degree at most n/k, a random k-element subset of the vertices contains an independent set of size at least $k/4$ with probability at least $1 - e^{-k/8}$.*

Proof. Let $v_1, \ldots, v_k$ be a sequence of k vertices chosen uniformly at random without replacement. Let I be the independent set formed greedily by, starting with the empty set, putting each v_i, sequentially as

$i = 1, 2, \ldots$, into I if doing so keeps I an independent set. During the process, when at most $k/4$ elements are added to I, the probability that a new v_i is added to I is at least $1 - \frac{(k/4)(n/k)}{n-k} \geq \frac{1}{2}$, since among the remaining $n - k$ vertices, at most $(k/4)(n/k)$ of them are adjacent to vertices already added to I at this point. It follows that $|I|$ stochastically dominates $\min\{X, k/4\}$, where X is distributed as Binomial$(k, 1/2)$. Thus $\mathbb{P}(|I| < k/4) \leq \mathbb{P}(X < k/4) \leq e^{-k/8}$ by the Chernoff bound. Therefore, $\{v_1, \ldots, v_k\}$ contains an independent set I of size at least $k/4$ with probability at least $1 - e^{-k/8}$. $\qquad\square$

We recall a basic result on VC dimension.

Theorem 3.3 (Sauer–Perles–Shelah theorem [26, 28, 32]). *If $\mathcal{S}$ is a set system on a ground set of n elements with VC dimension at most d, then $|\mathcal{S}| \leq \sum_{i=0}^{d} \binom{n}{i} \leq 2n^d$.*

Lemma 3.4. *Let $0 < \delta < 1$, and let m and d be positive integers. Let $\mathcal{S}$ be a δ-separated set system. Suppose that for a uniformly random m-element subset M, the restricted set system $\mathcal{S}|_M := \{T \cap M : T \in \mathcal{S}\}$ has VC dimension at most d with probability at least $3m^{2d}(1 - \delta)^m$. Then $|\mathcal{S}| \leq 2m^d$.*

Proof. Assume for contradiction that there exists such a set system with $|\mathcal{S}| = 2m^d + 1$. Let n be the size of the ground set. We have $|S \Delta T| \geq \delta n$ for all distinct $S, T \in \mathcal{S}$. Then, for each pair of distinct $S, T \in \mathcal{S}$, with probability at least $1 - (1 - \delta)^m$, M intersects $S \Delta T$, so that S and T remain distinct when restricted to M. Taking a union bound over all pairs of sets in $\mathcal{S}$, we see that with probability at least $1 - \binom{|\mathcal{S}|}{2}(1 - \delta)^m \geq 1 - 3m^{2d}(1 - \delta)^m$, all sets in $\mathcal{S}$ remain distinct when restricted to M, in which case VC-dim$(\mathcal{S}|_M) > d$ by Theorem 3.3 as $|\mathcal{S}| > 2m^d$, a contradiction to the hypothesis. $\qquad\square$

Lemma 3.5. *Let m and d be positive integers and $0 < \delta < 1$. Let G be a finite abelian group of order at least $24m^d$. Let X be a random $12m^d$-element subset of G, and Y a random m-element subset of G. If VC-dim$\{(A + x) \cap Y : x \in X\} \leq d$ with probability at least $e^{-m^d} + 3m^{2d}(1 - \delta)^m$, then $B = \{x : |A \Delta (A + x)| \leq \delta |G|\}$ has at least $|G|/(12m^d)$ elements.*

Proof. Suppose, on the contrary, that $|B| < |G|/(12m^d)$. Consider the Cayley graph on G generated by $B \setminus \{0\}$, i.e., there is edge between $x, y \in G$ whenenver $x - y \in B$. Applying Lemma 3.2 with $k = 12m^d$ to this graph, we find that with probability at least $1 - e^{-m^d}$, a random $12m^d$-element subset $X \subseteq G$ contains an independent set $I \subseteq X$ with $|I| \geq 3m^d$ with respect to this graph, i.e., $|(A + x) \Delta (A + y)| > \delta |G|$ for all distinct $x, y \in I$. It follows, by union bound and averaging, that we can fix such a set X so that VC-dim$\{(A + x) \cap Y : x \in X\} \leq d$ with probability at least $3m^{2d}(1 - \delta)^m$ for the random m-element set $Y \subseteq G$.

Note that $\{A + x : x \in I\}$ is a δ-separated set system with ground set G. Furthermore, VC-dim$\{(A + x) \cap Y : x \in I\} \leq$ VC-dim$\{(A + x) \cap Y : x \in X\} \leq d$ with probability at least $3m^{2d}(1 - \delta)^m$. So by Lemma 3.4, we have $|I| \leq 2m^d$, which contradicts the bound $|I| \geq 3m^d$ above. $\qquad\square$

Proof of Proposition 3.1. Let $0 < \delta < 1/2$. Consider $B = \{x : |A \Delta (A + x)| \leq \delta |G|\}$. Choose $m = C\delta^{-1} \log(1/\delta)$ where C is a sufficiently large constant. Then $e^{-m^d} < 1/20$ and $3m^{2d}(1 - \delta)^m < 2m^{2d} e^{-\delta m} < 1/20$.

If $|B| < |G|/(12m^d)$, then by Lemma 3.5, if X and Y are random $2m^d$-element subsets of G, then VC-dim$\{(A + x) \cap Y : x \in X\} > d$ with probability at least 0.9.

On the other hand, if $|B| \geq |G|/(12m^d)$, then by Lemma 2.4 there exists a subgroup H of G with $|H| \geq \delta^{o(1)}|B| \geq \delta^{d+o(1)}|H|$ such that $|A\Delta S| \leq \delta^{1-o(1)}|G|$ for some union S of H-cosets. By choosing $\delta = \varepsilon^{1+o(1)}$ so that $|A\Delta S| \leq \varepsilon|G|$, we obtain the desired result. $\qquad \square$

4 Removal lemma

In this section, we prove the removal lemma, Theorem 1.2, for bi-induced patterns.

The result is analogous to the induced removal lemma [2] which can be proved using a strong version of the graph regularity lemma. The usual way of proving the strong graph regularity lemma involves iteratively applying the graph regularity lemma. For our arithmetic setting, as we are concerned with bi-induced patterns, the situation is a bit easier: we simply apply the regularity lemma, Proposition 3.1, twice, where the second time we choose a smaller error parameter compared to the first time. If option (a) holds either time, then we can extract a bi-induced copy of F from each sample with high VC dimension. Otherwise, (b) holds, and we can modify A by a small amount to A', which must also have a bi-induced copy of F (since A is ε-far from bi-induced-F-free). The set A' is a union of H-cosets where H is a subgroup of bounded index, and we will show that a single bi-induced copy of F in A' leads to many copies.

Proof of Theorem 1.2. Let $V(F) = U \cup V$ be the vertex bipartition of F, where $|U| \geq |V|$. Let $d = |U| + \lceil \log_2 |U| \rceil$.

We may assume that $|G| \geq \varepsilon^{-\Omega(|V(F)|^2)}$ or else the conclusion is automatic from just a single bi-induced copy of F in A.

Suppose, for some $k = \varepsilon^{-O(|V(F)|)}$, with probability at least 0.9, random k-element subsets $X, Y \subseteq G$ satisfy VC-dim$\{(A+x) \cap Y : x \in X\} > d$, in which case there exist injective maps $U \to X$ and $V \to Y$ that bi-induce F in A by footnote 2. Then the probability that random injections $U \to G$ and $V \to G$ bi-induce F is at least $0.9\binom{k}{|U|}^{-1}\binom{k}{|V|}^{-1} \geq 0.9k^{-|U|-|V|} \geq \varepsilon^{O(|V(F)|^2)}$, since we can choose the random injection $U \to G$ by first choosing the random k-element subset $X \subset G$ and then taking a random injection $U \to X$, and similarly with V. With probability $1 - O_F(|G|^{-1})$ a random map $V(F) \to G$ is injective on U and V, so it bi-induces F with probability at least $\varepsilon^{O(|V(F)|^2)}$.

We apply Proposition 3.1 with two different parameters $\varepsilon_1 = \varepsilon/10$ and some ε_2 to be specified later. If option (a) is true in either case, then the previous paragraph implies the conclusion of the Theorem. Otherwise, we obtain subgroups H_1 and H_2 of G, such that for each $i \in \{1,2\}$, one has $h_i := |G|/|H_i| \leq \varepsilon_i^{-d-o(1)}$ and there exists some union S_i of H_i-cosets satisfying $|A\Delta S_i| \leq \varepsilon_i|G|$. Furthermore, we choose ε_2 so that $h_1\varepsilon_2|U||V| = 1/8$. In particular, $\varepsilon_2 \geq \varepsilon^{d+o(1)}$.

Let $H = H_1 \cap H_2$. So $|G|/|H| \leq h_1h_2 \leq \varepsilon^{-d^2-d-o(1)}$. We say that a coset $x+H$ of H is *good* if $|A\Delta(x+H)|/|H|$ is within $\eta := 1/(2|U||V|)$ of 0 or 1, and *bad* otherwise. At most an ε_2/η-fraction of H-cosets are bad, since otherwise bad H-cosets would together contribute more than $(\varepsilon_2/\eta)\eta|G|$ elements to $A\Delta S_2$ as S_2 is also a union of H-cosets, but this is impossible as $|A\Delta S_2| \leq \varepsilon_2|G|$.

Pick an arbitrary subgroup K of G containing exactly one element from each coset of H_1 (so that $G = H_1 \oplus K$ as a direct sum). Let $z \in H_1$ be chosen uniformly at random. Then $z + K + H$ is a union of $|K| = h_1$ many H-cosets. For each $y \in K$, the random H-coset $z + y + H$ is uniformly chosen from all

H-cosets in $y + H_1$. Applying the union bound, we see that the probability that $z + K + H$ contains a bad H-coset is at most $h_1 \varepsilon_2 / \eta < 2 h_1 \varepsilon_2 |U| |V| < 1/2$.

Let $A' \subseteq G$ be the union of H_1-cosets $y + H_1$, ranging over all $y \in K$ with $|A \cap (z + y + H)| \geq |H|/2$. Since A' and S_1 are both unions of H_1-cosets, we can apply linearity of expectation over H_1-cosets to deduce that $\mathbb{E}[|A' \Delta S_1|] \leq 2 |A \Delta S_1| \leq 2 \varepsilon_1 |G|$, and hence $\mathbb{E}[|A' \Delta A|] \leq \mathbb{E}[|A' \Delta S|] + |A \Delta S| \leq 3 \varepsilon_1 |G|$. Thus, with probability at least $1/2$, one has $|A' \Delta A| / |G| \leq 6 \varepsilon_1 < \varepsilon$.

Therefore there is some instance such that $|A' \Delta A| < \varepsilon |G|$, and $z + K + H$ is a union of good H-cosets.

Since A is ε-far from bi-induced-F-free, A' contains a bi-induced-copy of F. So there exist $x'_u, y'_v \in G$ over $u \in U$ and $v \in V$ such that for all $(u, v) \in U \times V$, one has $x'_u + y'_v \in A'$ if and only if $uv \in E(F)$. Since A' is a union of H_1-cosets, and there is an element of K in every H_1-coset, we may assume that $x'_u \in K$ for each $u \in U$ and $y'_v \in z + K$ for each $v \in V$.

Consider independent and uniform random elements $x_u \in x'_u + H$ for each $u \in U$, and $y_v \in y'_v + H$ for each $v \in V$. For each $(u, v) \in U \times V$, the random element $x_u + y_v$ is distributed uniformly in the H-coset $x'_u + y'_v + H$, which is a good H-coset since $x'_u + y'_v \in z + K$ as K is a subgroup. So with probability at least $1 - \eta$, one has $x_u + y_v \in A$ if and only if $x'_u + y'_v \in A'$, which in turn occurs if and only if $uv \in E(F)$. Taking a union bound over $(u, v) \in U \times V$, the following holds with probability at least $1 - |U| |V| \eta = 1/2$: for every $(u, v) \in U \times V$, one has $x'_u + y'_v \in A$ if and only if $uv \in E(F)$. Since each x_u and y_v is restricted to a single H-coset, it follows that a uniform random map $\phi : V(F) \to G$ bi-induces F with probability at least $\frac{1}{2}(|H| / |G|)^{|V(F)|} \geq \varepsilon^{(d^2 + d + o(1))|V(F)|}$. $\qquad\square$

5 Concluding remarks

We conjecture that the result can be extended to general groups, not necessarily abelian.

Conjecture 5.1. *Fix positive integers r and d. Let G be a group of exponent at most r, and $A \subseteq G$ a subset with VC dimension at most d. Then, for every $\varepsilon > 0$, there is a normal subgroup H of G of index at most $\varepsilon^{-O_{r,d}(1)}$ so that $|A \Delta S| \leq \varepsilon |G|$ for some union S of H-cosets.*

A special case of the conjecture, though with a somewhat stronger but non-quantitative conclusion, where one forbids a half-graph of fixed size (instead of assuming bounded VC dimension), was recently established by Conant, Pillay, and Terry [6] using model theoretic tools.

Note that the bounded exponent hypothesis in the conjecture above cannot be dropped. Indeed, if $G = \mathbb{Z}/p\mathbb{Z}$ with p prime, and $A = \{1, 2, \ldots, \lfloor p/2 \rfloor\}$, then VC-dim$A \leq 3$, while G has no non-trivial subgroups, so the conclusion of the conjecture is false. Nonetheless, there may be regularity lemmas using other structures in addition to subgroups. An example of such a result is discussed later in this section.

We also conjecture that the removal lemma should generalize to arbitrary groups as well, although it seems to be open even for the general abelian groups.

Conjecture 5.2. *Fix a bipartite graph F. Let G be a finite group. For every $0 < \varepsilon < 1/2$, if $A \subseteq G$ is ε-far from bi-induced-F-free, then the probability that a uniform random map $\phi : V(F) \to G$ bi-induces F is at least $\varepsilon^{O_F(1)}$.*

It seems likely that the theory developed by Breuillard, Green, and Tao [4, 5] on the structure of approximate groups should be useful in the case of nonabelian groups. We hope to study these problems in the future.

In classical results in additive combinatorics, such as Freiman's theorem, when the ambient group does not have many subgroups, generalized progressions and Bohr sets often play the role of subgroups when the group does not have many subgroups. For example, in Green and Ruzsa's [14] extension of Freiman's theorem to general abelian groups, the basic structural objects are *coset progressions*, which are sets of the form $P = Q + H$, where H is a subgroup, and Q is some generalized arithmetic progression $\{x_0 + i_1 x_1 + \cdots + i_d x_d : 0 \le i_j < \ell_j \text{ for each } j\}$, and the sum $Q + H$ is a direct sum in the sense that every element in $Q + H$ can be written as $q + h$ with $q \in Q$ and $h \in H$ in a unique way. We say that the progression is *proper* if all the terms $x_0 + i_1 x_1 + \cdots + i_d x_d$ in Q are distinct. We call d the *dimension* of the progression.

The Bogolyubov–Ruzsa lemma, Theorem 2.3, holds for general abelian groups (see [14, Section 5]; also see [24]).

Theorem 5.3 (Bogolyubov–Ruzsa lemma for general abelian groups)**.** *Let G be an abelian group, and $A \subseteq G$ a finite set such that $|A + A| \le K|A|$. Then $2A - 2A$ contains a proper coset progression P of dimension at most $d(K)$ and size at least $c(K)|A|$, for some constants $c(K), d(K) > 0$.*

By modifying the proof of Theorem 2.3 so that we apply Theorem 5.3 instead of 2.3, we obtain an analog of the first claim in Theorem 2.3 for general finite abelian groups. We are not sure if some variant of this result can be used to prove a removal lemma.

Proposition 5.4. *For every $\varepsilon > 0$ and $D = D(\varepsilon) \to \infty$ as $\varepsilon \to 0$, if G is a finite abelian group, and $A \subseteq G$ has VC dimension at most d, then there exist some proper coset progression P of dimension at most D and size $|P| \ge \varepsilon^{d+o(1)}|G|$, such that $|(A + x)\Delta A| \le \varepsilon|G|$ for all $x \in P$. Here $o(1)$ is some quantity that goes to zero as $\varepsilon \to 0$, at a rate depending on d and D.*

We conclude with the following related question that we do not know how to answer (even for $k = 2$). An affirmative answer would strengthen Szemerédi's theorem.

Question 5.5. Let k be a positive integer and $\delta > 0$. Let p be a sufficiently large prime, and $A \subseteq \mathbb{Z}/p\mathbb{Z}$ with $\delta p \le |A| \le (1 - \delta)p$. Can we always find a $2k$-term arithmetic progression in $\mathbb{Z}/p\mathbb{Z}$ where the first k terms lie in A and the last k terms lie outside of A?

If p had a small prime factor, then taking A to be a non-trivial subgroup of $\mathbb{Z}/p\mathbb{Z}$ gives a counterexample. To see the relevance to the rest of this paper, observe that such a $2k$-term arithmetic progression would bi-induce a half-graph on $2k$ vertices. For example, if $x - (k-1)d, x - (k-2)d, \ldots, x \in A$ and $x + d, \ldots, x + kd \notin A$, then $x_i = x - id$ and $y_j = jd$ have the property that, for $1 \le i, j \le d$, $x_i + y_j \in A$ if and only if $j \le i$.

References

[1] N. Alon, J. Balogh, B. Bollobás, and R. Morris, The structure of almost all graphs in a hereditary property, *J. Combin. Theory Ser. B* **101** (2011), 85–110. 2

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, *Combinatorica* 20 (2000), 451-476. 4, 9

[3] N. Alon, E. Fischer, and I. Newman, Efficient testing of bipartite graphs for forbidden induced subgraphs, *SIAM J. Comput.* **37** (2007), 959–976. 2

[4] E. Breuillard, B. Green, and T. Tao, Small doubling in groups. *Erdős centennial*, 129–151, Bolyai Soc. Math. Stud., 25, János Bolyai Math. Soc., Budapest, 2013. 11

[5] E. Breuillard, B. Green, and T. Tao, The structure of approximate groups, *Publ. Math. Inst. Hautes Études Sci.* **116** (2012), 115–221. 11

[6] G. Conant, A. Pillay and C. Terry, A group version of stable regularity, *Math. Proc. Camb. Philos. Soc.*, to appear. 3, 10

[7] D. Conlon and J. Fox, Bounds for graph regularity and removal lemmas, *Geom. Funct. Anal.* **22** (2012), 1191–1256. 4

[8] D. Conlon and J. Fox, Graph removal lemmas. *Surveys in combinatorics* **2013**, 1–49, London Math. Soc. Lecture Note Ser., 409, Cambridge Univ. Press, 2013. 4

[9] J. Fox, A new proof of the graph removal lemma, *Ann. of Math.* 174 (2011), 561–579. 4

[10] J. Fox, L. Gishboliner, A. Shapira, and R. Yuster, The removal lemma for tournaments, *J. Combin. Theory Ser. B*, to appear. 2

[11] J. Fox, J. Pach, and A. Suk, Erdős-Hajnal conjecture for graphs with bounded VC-dimension, to appear in *Discrete Comput. Geom.*, SoCG 2017 Special Issue. 2

[12] W. T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, *Geom. Funct. Anal.* **2** 1997, 322–337. 1

[13] B. Green, A Szemerédi-type regularity lemma in abelian groups, with applications, *Geom. Funct. Anal.* **15** (2005), 340–376. 3, 4

[14] B. J. Green and I. Z. Ruzsa, Freiman's theorem in an arbitrary abelian group, *J. London Math. Soc. (2)* **75** (2007), 163–175. 11

[15] D. Haussler, Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension, *J. Combin. Theory Ser. A*, **69** (1995), 217–232. 5

[16] M. Kneser, Abschätzungen der asymptotischen Dichte von Summenmengen, *Math. Zeitschr.* (in German) **58** (1953), 459–484. 6

[17] D. Král', O. Serra and L. Vena, A combinatorial proof of the removal lemma for groups, *J. Combin. Theory Ser. A* **116** (2009), 971–978. 4

[18] D. Král', O. Serra and L. Vena, A removal lemma for systems of linear equations over finite fields, *Israel J. Math.* **187** (2012), 193–207. 4

[19] L. Lovász and B. Szegedy, Regularity partitions and the topology of graphons, An Irregular Mind, Imre Bárány, József Solymosi, and Gábor Sági editors, *Bolyai Society Mathematical Studies* **21** (2010), 415–446. 2

[20] M. Malliaris and S. Shelah, Regularity lemmas for stable graphs, *Trans. Amer. Math. Soc.* **366** (2014), 1551–1585. 2

[21] I. Z. Ruzsa, Generalized arithmetical progressions and sumsets, *Acta Math. Hungar.* **65** (1994), 379–388. 5

[22] I. Z. Ruzsa, An analog of Freiman's theorem in groups, *Astérisque* **258** (1999), 323–326. 6

[23] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, in Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai 18, Volume II, 939–945. 4

[24] T. Sanders, On the Bogolyubov-Ruzsa lemma, *Anal. PDE* **5** (2012), no. 3, 627–655. 5, 11

[25] T. Sanders, The structure theory of set addition revisited, *Bull. Amer. Math. Soc.* **50** (2013), 93–127. 5

[26] N. Sauer, On the density of families of sets, *J. Combinatorial Theory Ser. A* **13** (1972), 145–147. 8

[27] A. Shapira, A proof of Green's conjecture regarding the removal properties of sets of linear equations, *J. London Math. Soc.* **81** (2010), 355–373. 4

[28] S. Shelah, A combinatorial problem; stability and order for models and theories in infinitary languages, *Pacific J. Math.* **41** (1972), 247–261. 8

[29] T. C. Tao and H. V. Vu., *Additive combinatorics*, Cambridge University Press, 2006. 6

[30] C. Terry and J. Wolf, Stable arithmetic regularity in the finite-field model, *Bull. Lond. Math. Soc.* **51** (2019), 70–88. 3

[31] E. Szemerédi, Regular partitions of graphs, Problémes combinatoires et th eorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloq. Internat. CNRS, vol. 260, CNRS, Paris, 1978, pp. 399–401. 1

[32] V. N. Vapnik and A. Ja. Červonenkis, The uniform convergence of frequencies of the appearance of events to their probabilities (Russian), *Teor. Verojatnost. i Primenen.* **16** 1971, 264–279. 8

AUTHORS

Noga Alon
Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv 69978, Israel
and
Department of Mathematics, Princeton University, Princeton, NJ 08544, USA
nogaa [at] tau [dot] ac [dot] il
https://web.math.princeton.edu/~nalon/

Jacob Fox
Department of Mathematics, Stanford University, Stanford, CA 94305, USA
jacobfox [at] stanford [dot] edu
http://stanford.edu/~jacobfox/

Yufei Zhao
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
yufeiz [at] mit [dot] edu
http://yufeizhao.com