

MIT Open Access Articles

Testing Linear-Invariant Properties

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Tidor, Jonathan and Zhao, Yufei. 2022. "Testing Linear-Invariant Properties." SIAM Journal on Computing, 51 (4).

As Published: 10.1137/21m1397246

Publisher: Society for Industrial & Applied Mathematics (SIAM)

Persistent URL: <https://hdl.handle.net/1721.1/145888>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



TESTING LINEAR-INVARIANT PROPERTIES*

JONATHAN TIDOR[†] AND YUFEI ZHAO[†]

Abstract. We study the property testing of functions $\mathbb{F}_p^n \rightarrow [R]$ for fixed prime p and positive integer R . We work in the natural model where we are allowed to query the function on a random subspace of constant dimension. We say that a property is testable if queries of this form can detect the property with one-sided error. Furthermore, a property is proximity oblivious-testable (PO-testable) if the test is also independent of the proximity parameter ϵ . It is known that a number of natural properties such as linearity and being a low degree polynomial are PO-testable. These properties are examples of linear-invariant properties, meaning that they are preserved under linear automorphisms of the domain. Following work of Kaufman and Sudan, the study of linear-invariant properties has been an important problem in arithmetic property testing. A central conjecture in this field, proposed by Bhattacharyya, Grigorescu, and Shapira, is that a linear-invariant property is testable if and only if it is semi-subspace-hereditary. We prove two results; the first resolves this conjecture and the second classifies PO-testable properties: (1) A linear-invariant property is testable if and only if it is semi-subspace-hereditary. (2) A linear-invariant property is PO-testable if and only if it is locally characterized. Our innovations are twofold. We give a more powerful version of the compactness argument first introduced by Alon and Shapira. This relies on a new strong arithmetic regularity lemma in which one mixes different levels of Gowers uniformity. This allows us to extend the work of Bhattacharyya, Fischer, Hatami, Hatami, and Lovett by removing the bounded complexity restriction in their work. Our second innovation is a novel recoloring technique called patching that builds on earlier work by the authors and Fox. This Ramsey-theoretic technique is critical for working in the linear-invariant setting and allows us to remove the translation-invariant restriction present in previous work.

Key words. property testing, higher-order Fourier analysis, sublinear time algorithms

MSC codes. 11B30, 68Q25, 68W20

DOI. 10.1137/21M1397246

1. Introduction. In property testing, the aim is to find randomized algorithms that distinguish objects that have some property from those that are far from satisfying the property by querying the given large object at a small number of locations. Property testing emerged from the linearity test of Blum, Luby, and Rubinfeld [10] and was formally defined and systematically studied by Rubinfeld and Sudan [30] and Goldreich, Goldwasser, and Ron [14]. There have been important developments especially in the following two settings: graph property testing and arithmetic property testing.

Two representative problems are (1) given a large graph, test whether the graph is triangle-free or ϵ -far from triangle-free (an n -vertex graph is ϵ -far from a graph property if one needs to add and/or remove more than ϵn^2 edges in order to satisfy the property), and (2) given a function $f: \mathbb{F}_p^n \rightarrow \mathbb{F}_p$, test whether f is linear or ϵ -far from linear (a function is ϵ -far from an arithmetic property if one needs to change the value of the function on more than an ϵ -fraction of the domain in order to satisfy the property). In both cases, it is known that one can achieve the desired goal by

*Received by the editors February 25, 2021; accepted for publication (in revised form) April 18, 2022; published electronically July 21, 2022.

<https://doi.org/10.1137/21M1397246>

Funding: The work of the first author was supported by NSF Graduate Research Fellowship Program DGE-1745302. The work of the second author was supported by NSF award DMS-1764176, the MIT Solomon Buchsbaum Fund, and a Sloan Research Fellowship.

[†]Massachusetts Institute of Technology, Cambridge, MA 02139 USA (jtidor@mit.edu, yufeiz@mit.edu).

sampling a fixed number of entries repeatedly $C(\epsilon)$ times. For testing whether a graph is triangle-free [31], one samples a uniform random triple of vertices and checks whether they form a triangle, and for testing linearity [10], one samples $x, y \in \mathbb{F}_p^n$ uniformly and checks if $f(x) + f(y) = f(x + y)$.

In this paper we give a property testing algorithm for a very general class of arithmetic properties. The goal is to determine whether a function $f: \mathbb{F}_p^n \rightarrow [R] := \{1, \dots, R\}$ (with fixed prime p and positive integer R) satisfies some given property or is ϵ -far from satisfying the property. All the properties we consider are *linear-invariant* in the sense that they are invariant under automorphisms of the domain vector space.¹ Linear-invariant properties form an important general class of arithmetic properties, e.g., the work of Kaufman and Sudan [26] “highlights linear-invariance as a central theme in algebraic property testing.”

An oblivious tester is one which works as follows: the tester produces a positive integer $d = d(\epsilon)$ and an oracle provides the tester with the restriction $f|_U$ where U is a uniform random d -dimensional linear subspace of the domain (if the domain is large enough that such a subspace exists; if the domain has dimension strictly less than d , the oracle provides the tester with all of f). We say that a property \mathcal{P} is *testable* if there exists an oblivious tester with one-sided error (and constant query-complexity) for the property. That is, we require our tester to accept functions f satisfying \mathcal{P} with probability 1 and reject functions that are ϵ -far from satisfying \mathcal{P} with probability at least $\delta = \delta(\epsilon)$ for some function $\delta: (0, 1) \rightarrow (0, 1)$. Furthermore, we say that \mathcal{P} is *proximity oblivious-testable* (PO-testable) if $d = d(\epsilon)$ is a constant independent of ϵ . The idea of PO-testability was introduced by Goldreich and Ron [15], who, among other results, classified the PO-testable graph properties.

One surprising feature of property testing is that many natural properties, such as linearity, are testable and even PO-testable. A key feature of linearity is that it is subspace-hereditary, meaning that if $f: \mathbb{F}_p^n \rightarrow \mathbb{F}_p$ is linear, then the same is true for $f|_U$ for every linear subspace $U \leq \mathbb{F}_p^n$. To be precise, we say that a linear-invariant property \mathcal{P} is *subspace-hereditary* if for every $f: \mathbb{F}_p^n \rightarrow [R]$ satisfying \mathcal{P} and every linear subspace $U \leq \mathbb{F}_p^n$, the restriction $f|_U$ also satisfies \mathcal{P} .

A central conjecture in this field, first proposed by Bhattacharyya, Grigorescu, and Shapira, is that all linear-invariant, subspace-hereditary properties are testable [9, Conjecture 4]. In fact, they conjecture that the slightly larger class of *semi-subspace-hereditary* properties are testable and prove that no other properties can be tested.

DEFINITION 1.1. A linear-invariant property \mathcal{P} is semi-subspace-hereditary if there exists a subspace-hereditary property \mathcal{Q} such that

- (i) every function satisfying \mathcal{P} also satisfies \mathcal{Q} ;
- (ii) for all $\epsilon > 0$, there exists $N(\epsilon)$ such that if $f: \mathbb{F}_p^n \rightarrow [R]$ satisfies \mathcal{Q} and is ϵ -far from satisfying \mathcal{P} , then $n < N(\epsilon)$.

It is known that there are subspace-hereditary properties where the dimension d sampled must grow as the proximity parameter ϵ approaches 0. To be PO-testable, a property must satisfy the following more restrictive condition.

DEFINITION 1.2. A linear-invariant property \mathcal{P} is locally characterized if there exists some d such that the following holds. For every $f: \mathbb{F}_p^n \rightarrow [R]$ with $n \geq d$, the function f satisfies \mathcal{P} if and only if $f|_U$ satisfies \mathcal{P} for every $U \leq \mathbb{F}_p^n$ of dimension d .

¹It would also be possible to consider this problem with base field \mathbb{F}_q for a prime power q . While many of our techniques extend to this setting, there are also many increased technical difficulties, so we do not attempt to consider this problem here.

Our first result is a resolution of the conjecture of Bhattacharyya, Grigorescu, and Shapira, classifying the testable linear-invariant properties. Our second result is a classification of the PO-testable linear-invariant properties.

THEOREM 1.3. *A linear-invariant property is testable if and only if it is semi-subspace-hereditary.*

THEOREM 1.4. *A linear-invariant property is PO-testable if and only if it is locally characterized.*

Remark 1.5. Note that under our definition of an oblivious tester, the tester does not know the dimension of the domain. This rules out some “unnatural” properties such as those properties that behave differently depending on whether the dimension of the domain is even or odd.

Previous work in arithmetic property testing has focused on a number of special cases including monotone properties [27, 32], “complexity 1” properties over \mathbb{F}_2 [9], and bounded complexity translation-invariant properties [7].

We note that very little was previously known about general linear-invariant properties. One simple way to define a class of linear-invariant properties can be done by, for example, choosing an arbitrary subset of “allowable” maps $\mathbb{F}_p^2 \rightarrow [R]$ and defining a property of functions $\mathbb{F}_p^n \rightarrow [R]$ to consist of those whose restriction to every two-dimensional linear subspace is allowable. Even this class of two-dimensionally defined patterns was not known to be testable in general prior to this work.

Our innovations are twofold. We prove a strong arithmetic regularity lemma which, unlike previous arithmetic regularity lemmas, mixes different levels of Gowers uniformity. This allows us to give a more powerful version of the *compactness argument* first introduced by Alon and Shapira [4]. With this tool we can remove the *bounded complexity* restriction that was present in all previous work.

Our second innovation is a novel recoloring technique we call *patching*. This technique is critical for working in the linear-invariant setting and allows us to handle an important obstacle encountered by previous works.

In the rest of this section we give a summary of the proof of the main theorem and its relation to previous work.

1.1. Graph removal lemmas and property testing. We begin with an overview of graph removal lemmas and their proof techniques (see the survey [11] for a more detailed discussion).

The foundational result in this field is the triangle removal lemma of Ruzsa and Szemerédi [31]. This result states that for all $\epsilon > 0$ there exists $\delta > 0$ such that any n -vertex graph with at most δn^3 triangles can be made triangle-free by removing ϵn^2 edges. It is immediate from the definitions that this result implies that triangle-freeness is testable (and in fact PO-testable). In general we will work with such *removal lemmas* and deduce corresponding property testing results from them. The triangle removal lemma was generalized to the graph removal lemma, first stated explicitly by Alon et al. [1] and by Füredi [13].

A key tool for proving the graph removal lemma is a regularity lemma, namely Szemerédi’s graph regularity lemma. Roughly speaking, the proof proceeds by using this regularity lemma to partition the input graph G into a small number of structured components. Then we “clean up” G by removing at most ϵn^2 edges. This is done in such a way that either the resulting graph is H -free or the original graph G contains many copies of H .

An important extension of the graph removal lemma is the induced graph removal lemma, proved by Alon et al. [2]. The induced graph removal lemma states that for every graph H , for all $\epsilon > 0$ there exists $\delta > 0$ such that every n -vertex graph with at most $\delta n^{v(H)}$ induced copies of H can be made induced H -free by adding and/or removing at most ϵn^2 edges (here *induced H -free* means not containing any induced subgraph isomorphic to H).

The original proof of the induced graph removal lemma relies on an extension of Szemerédi's graph regularity lemma known as the "strong regularity lemma." Using such a regularity lemma combined with a random sampling argument, one can produce a "regular model," that is, a large induced subgraph $X := G[U]$ (on a constant fraction of the vertices of G) that is very regular and approximates the original graph well in a certain sense. Then we "clean up" G by adding and/or removing at most ϵn^2 edges in such a way that if the resulting graph is not induced H -free, then X (in the original graph) must contain many induced copies of H .

Both the graph removal lemma and the induced graph removal lemma can easily be extended to remove any finite collection \mathcal{H} of graphs. Alon and Shapira [4] extended the induced graph removal lemma to an infinite collection of graphs. Namely they prove that for a (possibly infinite) set \mathcal{H} of graphs and for $\epsilon > 0$ there exist $\delta > 0$ and k such that the following holds: if G is an n -vertex graph with at most $\delta n^{v(H)}$ copies of H for all $H \in \mathcal{H}$ with k or fewer vertices, then G can be made induced \mathcal{H} -free by adding and/or removing at most ϵn^2 edges (meaning the modified graph has no induced subgraph isomorphic to H for every $H \in \mathcal{H}$). Despite the strange statement of this result (the hypothesis that G has low H -density only for the small graphs $H \in \mathcal{H}$) this theorem immediately implies (and is equivalent to) the fact that every hereditary graph property is testable with constant query-complexity and one-sided error.

This series of works, in addition to being important results in their own right, gives a framework for proving constant query-complexity property testing algorithms in other settings, given an appropriate regularity lemma. For example, a hypergraph regularity lemma is known, proven by Gowers [17] and independently by Rödl et al. [28]. Using the above framework, one can use the hypergraph regularity lemma to deduce an infinite induced hypergraph removal lemma [29]. Consequently, every hereditary hypergraph property is testable with constant query-complexity and one-sided error. We will be interested in seeing if this framework, combined with an appropriate arithmetic regularity lemma, can prove an infinite induced arithmetic removal lemma.

1.2. Arithmetic analogues. The problem of property testing for functions $f: \mathbb{F}_p^n \rightarrow [R]$ has been intensively studied, starting with the classic work of Blum, Luby, and Rubinfeld [10] on linearity testing. Much of the work focuses on testing whether some function $f: \mathbb{F}_p^n \rightarrow \mathbb{F}_p$ has certain algebraic properties (e.g., a polynomial of some given type) [3, 26]. There is also much interest in testing properties that do not arrive from algebraic characterizations. Below we give an overview of the developments related to property testing in \mathbb{F}_p^n from a perspective that is parallel to the graph regularity method developments discussed earlier.

The first arithmetic regularity lemma was proved by Green [19] using Fourier-analytic techniques, and it laid the groundwork for further developments of the regularity method in the arithmetic setting. These regularity lemmas have since found many applications in additive combinatorics and related fields. In particular, combined with the graph removal framework described above, Green's regularity lemma is

suitable for proving an arithmetic removal lemma for “complexity 1” systems of linear forms (see section 3 for the definition of complexity; roughly speaking, a system of linear forms is complexity 1 if it can be controlled by Fourier-analytic means); e.g., see [6].

Král’, Sera, and Vena [27] and independently Shapira [32] bypass the need for an arithmetic regularity lemma and prove the full arithmetic removal lemma by a direct reduction from the hypergraph removal lemma. Their results imply that all linear-invariant, subspace-hereditary *monotone* properties are testable with constant query-complexity and one-sided error. (A property of functions $\mathbb{F}_p^n \rightarrow \{0, 1\}$ is *monotone* if changing 1’s to 0’s preserves the property.)

Note that the above result is an arithmetic removal lemma and not an *induced* arithmetic removal lemma (hence the restriction to monotone properties). Due to the nature of the reduction, the techniques do not seem to be capable of deducing the induced arithmetic removal lemma from the induced hypergraph removal lemma.

An alternative approach is to apply the strong graph regularity approach [2] of proving the induced graph removal lemma to Green’s arithmetic regularity lemma. However, there is also a major obstacle to the approach, related to the fact that the origin plays a special role in a vector space while there is no corresponding feature of graphs. It turns out that it is not always possible to regularize the space in a neighborhood of the origin [20].

Bhattacharyya, Grigorescu, and Shapira [9] managed to overcome this obstacle in the special case of vector spaces over \mathbb{F}_2 . They follow the above strategy, implementing the strong regularity idea [2] in the style of Green’s arithmetic regularity [19] along with one additional tool, namely a Ramsey-theoretic result, to prove an infinite induced arithmetic removal lemma for complexity 1 patterns over \mathbb{F}_2 . Unfortunately, it is known [20] that this Ramsey-theoretic result fails over all finite fields other than \mathbb{F}_2 .

Bhattacharyya, Fischer, and Lovett [8] managed to overcome this obstacle in a different special case, namely for translation-invariant patterns. When all patterns considered are translation-invariant, the origin no longer plays a special role and one can essentially ignore it while carrying out the strong regularity framework. In addition, [8] allows one to handle higher complexity patterns, which requires developing and applying tools from higher-order Fourier analysis.

Higher-order Fourier analysis plays a central role in modern additive combinatorics. These techniques were initiated by Gowers [16] in his celebrated new proof of Szemerédi’s theorem, and further developed in a sequence of works by Green, Tao, and Ziegler [21, 22, 23] settling classical conjectures on the asymptotics of prime numbers patterns. A parallel theory of higher-order Fourier analysis was developed in finite field vector spaces by Bergelson, Tao, and Ziegler [5, 33, 34], leading to an inverse Gowers theorem over finite fields vector spaces.

For applications to property testing, this line of work culminated in the work of Bhattacharyya et al. [7] (extending [8]), who applied the inverse Gowers theorem over finite fields and developed further equidistribution tools to prove an infinite induced arithmetic removal lemma for all linear-invariant, subspace-hereditary properties that are also translation-invariant and bounded complexity. Their work follows the strong regularity framework of [2, 4]. Our results improve upon this work by removing the translation-invariant and bounded complexity restrictions.

In addition to their property testing algorithm, Bhattacharyya et al. [7] proved that a large class of somewhat algebraically structured properties are indeed affine-invariant, subspace-hereditary, and locally characterized. These are the so-called degree-structural properties. A simple extension of their result [25, Theorem 16.3]

implies that the larger class of “homogeneous degree-structural properties” is linear-invariant, linear subspace-hereditary (but not affine-invariant and not subspace-hereditary), and locally characterized, and thus these properties are testable by our main theorem. As an example, one can test whether a function $\mathbb{F}_p^n \rightarrow \mathbb{F}_p$ can be written as $A^2 + B^2$ where both A and B are homogeneous polynomials of some given degree d .

1.3. Our contributions.

1.3.1. Patching. In this paper, building on the authors’ earlier work with Fox [12] for complexity 1 patterns, we develop a new technique called “patching” that allows us to overcome the obstacle faced by earlier approaches, namely that a neighborhood of the origin cannot be regularized and fails certain Ramsey properties (unless working over \mathbb{F}_2). In essence, the patching result states that if there exists some map $f: \mathbb{F}_p^n \rightarrow [R]$ that has low density of some colored patterns \mathcal{H} for n large enough, then for all m there must exist some map $g: \mathbb{F}_p^m \rightarrow [R]$ that has no \mathcal{H} -instances. (Informally, a colored pattern is a linear configuration with specified colors of the points.)

THEOREM 1.6 (informal patching result). *For every set of colored patterns \mathcal{H} , there exist $\epsilon_0 > 0$ and n_0 such that the following holds. Either*

- *for every n , there exists a function $f: \mathbb{F}_p^n \rightarrow [R]$ that is \mathcal{H} -free, or*
- *for every function $f: \mathbb{F}_p^n \rightarrow [R]$ with $n \geq n_0$, the H -density in f is at least ϵ_0 for some $H \in \mathcal{H}$.*

Our proof proceeds in two steps. First, as in [7], following the strong regularity framework of [2] for proving induced graph removal lemmas, we apply a strong arithmetic regularity lemma, which produces a partition \mathfrak{B} of \mathbb{F}_p^n and a “regular model” $X \subseteq \mathbb{F}_p^n$ made up of a randomly sampled set of atoms from \mathfrak{B} . Unlike in the graph setting, we cannot ensure that the map $f: \mathbb{F}_p^n \rightarrow [R]$ is very regular on every atom of $\mathfrak{B}|_X$. In particular, it may be impossible to guarantee that f is regular on the atom containing the origin. Instead we only ensure that almost every atom of X is very regular. Unlike earlier proofs of removal lemmas, our “recoloring algorithm” has two components: for the regular atoms we “clean up” f as usual, while for the irregular atoms we apply our patching result. Our patching result implies that there is some new global coloring $g: \mathbb{F}_p^n \rightarrow [R]$ that avoids some appropriate set of colored patterns. To complete the proof we “patch” f by replacing it by g on all of the irregular atoms. If f has low density of some set of colored pattern, then our argument shows that these patterns cannot appear in the recoloring, thereby completing the proof of the induced arithmetic removal lemma. This patching result is stronger than the corresponding result in [12] since for this application the patching must also respect the structure of the partition given by \mathfrak{B} .

Our proof does not give effective bounds on the rejection probability function $\delta(\epsilon)$ guaranteed by Theorem 1.3. The ineffectiveness is due to the fact that the current best-known bounds on the inverse theorem for nonclassical polynomials are ineffective (the same occurs in [7]).

1.3.2. Unbounded complexity. The technique used to handle infinite removal lemmas is a compactness argument first introduced by Alon and Shapira [4] in the graph setting. A key ingredient of their proof is a strong regularity lemma.

Bhattacharyya et al. [7] prove that all linear-invariant subspace-hereditary properties that are also translation-invariant and bounded complexity are testable. Their

result follows from an infinite removal lemma for arithmetic patterns of bounded complexity. The proof of this result involves a strong arithmetic regularity lemma and a compactness argument in the spirit of Alon and Shapira.

To remove the bounded complexity assumption from [7], we prove a new strong arithmetic regularity lemma obtained by iterating a weaker arithmetic regularity lemma. The key innovation here is that the level of Gowers uniformity used in each iteration is allowed to increase at each step of the process.

2. Colored patterns and removal lemmas. Theorems 1.3 and 1.4 both follow from an arithmetic removal lemma for colored linear patterns. In this section we define these objects and state the main removal lemma.

DEFINITION 2.1. A linear form over \mathbb{F}_p in ℓ variables is an expression L of the form

$$L(x_1, \dots, x_\ell) = \sum_{i=1}^{\ell} c_i x_i$$

with $c_i \in \mathbb{F}_p$. For any \mathbb{F}_p -vector space V , the linear form L gives rise to a function $L: V^\ell \rightarrow V$ that is linear in each variable.

DEFINITION 2.2. For a prime p and a finite set \mathcal{S} , an \mathcal{S} -colored pattern over \mathbb{F}_p consisting of m linear forms in ℓ variables is a pair (\mathbf{L}, ψ) given by a system $\mathbf{L} = (L_1, \dots, L_m)$ of m linear forms in ℓ variables and a coloring $\psi: [m] \rightarrow \mathcal{S}$. Given a finite-dimensional \mathbb{F}_p -vector space V and a function $f: V \rightarrow \mathcal{S}$, an (\mathbf{L}, ψ) -instance in f is some $\mathbf{x} \in V^\ell$ such that $f(L_i(\mathbf{x})) = \psi(i)$ for all $i \in [m]$. An instance is called generic if x_1, \dots, x_ℓ are linearly independent. We say that (\mathbf{L}, ψ) is translation-invariant if the coefficient of x_1 is 1 in each of L_1, \dots, L_m .

Given a finite-dimensional \mathbb{F}_p -vector space V and functions $f_1, \dots, f_m: V \rightarrow [-1, 1]$, we write

$$\Lambda_{\mathbf{L}}(f_1, \dots, f_m) := \mathbb{E}_{\mathbf{x} \in V^k} [f_1(L_1(\mathbf{x})) \cdots f_m(L_m(\mathbf{x}))].$$

DEFINITION 2.3. For an \mathcal{S} -colored pattern over \mathbb{F}_p consisting of m linear forms in k variables (\mathbf{L}, ψ) , a finite-dimensional \mathbb{F}_p -vector space V , and a function $f: V \rightarrow \mathcal{S}$, define the (\mathbf{L}, ψ) -density in f to be $\Lambda_{\mathbf{L}}(f_1, \dots, f_m)$ where $f_i := 1_{f^{-1}(\psi(i))}$ for each $i \in [m]$.

Our main removal lemma is the following result.

THEOREM 2.4 (main removal lemma). Fix a prime p and a finite set \mathcal{S} . Let \mathcal{H} be a (possibly infinite) set of \mathcal{S} -colored patterns over \mathbb{F}_p . For every $\epsilon > 0$, there exists a finite set $\mathcal{H}_\epsilon \subseteq \mathcal{H}$ and $\delta = \delta(\epsilon, \mathcal{H}) > 0$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space. If $f: V \rightarrow \mathcal{S}$ has H -density at most δ for every $H \in \mathcal{H}_\epsilon$, then there exists a recoloring $g: V \rightarrow \mathcal{S}$ that agrees with f on all but an at most ϵ -fraction of V such that g has no generic H -instances for every $H \in \mathcal{H}$.

There are several difficulties in the proof of the main removal lemma. The first is that individual patterns $H \in \mathcal{H}$ may have “infinite complexity.” Second, the set of patterns \mathcal{H} may be infinite. Complicating this, even if all patterns in \mathcal{H} have finite complexity, these complexities can be unbounded. Finally, there are major difficulties related to the fact that the patterns in \mathcal{H} are not necessarily translation-invariant.

We use a trick we call “projectivization” to reduce to the case where all patterns have finite complexity.

A “compactness argument” due to Alon and Shapira [4] reduces the problem of an infinite collection of patterns to a finite one at the expense of requiring a stronger arithmetic regularity lemma. If the collection of patterns is all complexity at most d , we only require a strong U^{d+1} -regularity lemma with rapidly decreasing error parameter. In the most general case when the collection of patterns has unbounded complexity we require an even stronger regularity lemma where the error parameter rapidly decreases and the degree of the uniformity norm rapidly increases.

Unless we restrict to the special case where all patterns in \mathcal{H} are translation-invariant, the origin of the vector space plays a special role. This is unfortunate because it is impossible to regularize a function in the neighborhood of the origin. Since regularity methods are useless here, we turn to a new technique called patching, originally introduced by the authors and Fox [12], to deal with the portions of the vector space that cannot be regularized.

We now explain the “projectivization” trick. This will allow us to reduce to the case where all patterns have finite complexity at the expense of requiring that the functions considered are “projective.” This restriction can be intuitively thought of as requiring that the functions live on projective space instead of affine space. We first define what it means for a function to be projective and then state the projective removal lemma, Theorem 2.8, from which we will later deduce Theorem 2.4.

DEFINITION 2.5. Let \mathcal{S} be a finite set equipped with a group action of \mathbb{F}_p^\times that we denote $c \cdot s$ for $c \in \mathbb{F}_p^\times$ and $s \in \mathcal{S}$. Given a finite-dimensional \mathbb{F}_p -vector space V , a function $f: V \rightarrow \mathcal{S}$ is projective if it preserves the action of \mathbb{F}_p^\times , i.e., $f(cx) = c \cdot f(x)$ for all $c \in \mathbb{F}_p^\times$ and all $x \in V$.

We give one example of a class of projective functions which will be the one we need at the end of the argument. For a finite set \mathcal{S} , define $\bar{\mathcal{S}} := \mathcal{S}^{\mathbb{F}_p^\times}$ with \mathbb{F}_p^\times -action defined by

$$b' \cdot (c_b)_{b \in \mathbb{F}_p^\times} := (c_{b'b})_{b \in \mathbb{F}_p^\times}.$$

Then for any function $f: V \rightarrow \mathcal{S}$, one can create a projective function $\bar{f}: V \rightarrow \bar{\mathcal{S}}$ as

$$\bar{f}(x) := (f(bx))_{b \in \mathbb{F}_p^\times}.$$

DEFINITION 2.6. A list of linear forms $\mathbf{L} = (L_1, \dots, L_m)$ is finite complexity if no form is identically equal to zero, i.e., $L_i \not\equiv 0$ for all $i \in [m]$, and no two forms are linearly dependent, i.e., $L_i \not\equiv cL_j$ for all $i \neq j$ and $c \in \mathbb{F}_p$. (Later we will define the complexity of a system of linear forms to be a nonnegative integer or infinity.)

DEFINITION 2.7. Fix a prime p and a positive integer ℓ . We consider two particular systems of linear forms. For $\mathbf{i} = (i_1, \dots, i_\ell) \in \mathbb{F}_p^\ell$, define

$$L_{\mathbf{i}}^\ell(x_1, \dots, x_\ell) := i_1x_1 + \dots + i_\ell x_\ell.$$

Then define $\mathbf{L}^\ell := (L_{\mathbf{i}}^\ell)_{\mathbf{i} \in \mathbb{F}_p^\ell}$, the system of p^ℓ linear forms in ℓ variables that defines an ℓ -dimensional subspace.

Let $E_\ell \subset \mathbb{F}_p^\ell$ be the set of nonzero vectors whose first nonzero coordinate is 1. Then define $\bar{\mathbf{L}}^\ell := (L_{\mathbf{i}}^\ell)_{\mathbf{i} \in E_\ell}$, a system of $(p^\ell - 1)/(p - 1)$ linear forms in ℓ variables.

Note that unlike \mathbf{L}^ℓ , the system $\bar{\mathbf{L}}^\ell$ has finite complexity. For technical reasons, it will be convenient to reduce the removal lemma for general patterns to the case where all patterns are defined by a system of the form \mathbf{L}^ℓ . Then we reduce this to

the following projective removal lemma where all patterns are defined by a system of the form $\overline{\mathbf{L}}^\ell$.

THEOREM 2.8 (projective removal lemma). *Fix a prime p and a finite set \mathcal{S} equipped with an \mathbb{F}_p^\times -action. Let \mathcal{H} be a (possibly infinite) set consisting of \mathcal{S} -colored patterns over \mathbb{F}_p of the form $(\overline{\mathbf{L}}^\ell, \psi)$ where ℓ is some positive integer and $\psi: E_\ell \rightarrow \mathcal{S}$ is some map (see Definition 2.7 for the definition of $\overline{\mathbf{L}}^\ell$ and E_ℓ). For every $\epsilon > 0$, there exists a finite subset $\mathcal{H}_\epsilon \subseteq \mathcal{H}$ and $\delta = \delta(\epsilon, \mathcal{H}) > 0$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space. If $f: V \rightarrow \mathcal{S}$ is a projective function with H -density at most δ for every $H \in \mathcal{H}_\epsilon$, then there exists a projective recoloring $g: V \rightarrow \mathcal{S}$ that agrees with f on all but an at most ϵ -fraction of V such that g has no generic H -instances for every $H \in \mathcal{H}$.*

3. Preliminaries on higher-order Fourier analysis.

3.1. Gowers norms and complexity. Our main tool for studying arithmetic patterns is the Gowers uniformity norms. Informally, if a colored pattern H has complexity d , then any coloring which is “quasirandom with respect to the Gowers U^{d+1} -norm” will have close to the expected number of copies of H .

DEFINITION 3.1. *Fix a prime p , a finite-dimensional \mathbb{F}_p -vector space V , and an abelian group G . Given a function $f: V \rightarrow G$ and a shift $h \in V$, define the additive derivative $D_h f: V \rightarrow G$ by*

$$(D_h f)(x) := f(x+h) - f(x).$$

Given a function $f: V \rightarrow \mathbb{C}$ and a shift $h \in V$, define the multiplicative derivative $\Delta_h f: V \rightarrow \mathbb{C}$ by

$$(\Delta_h f)(x) := f(x+h)\overline{f(x)}.$$

DEFINITION 3.2. *Fix a prime p and a finite-dimensional \mathbb{F}_p -vector space V . Given a function $f: V \rightarrow \mathbb{C}$ and $d \geq 1$, the Gowers uniformity norm $\|f\|_{U^d}$ is defined by*

$$\|f\|_{U^d} := |\mathbb{E}_{x, h_1, \dots, h_d \in V} (\Delta_{h_1} \cdots \Delta_{h_d} f)(x)|^{1/2^d}.$$

DEFINITION 3.3. *A system $\mathbf{L} = (L_1, \dots, L_m)$ of m linear forms in ℓ variables is complexity at most d if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $f_1, \dots, f_\ell: V \rightarrow [-1, 1]$ it holds that*

$$(3.1) \quad |\Lambda_{\mathbf{L}}(f_1, \dots, f_\ell)| \leq \epsilon \quad \text{whenever} \quad \min_{1 \leq i \leq \ell} \|f_i\|_{U^{d+1}} \leq \delta.$$

The complexity of \mathbf{L} is the smallest d such that the above holds, and infinite otherwise.

(3.1) is known as the counting lemma and will be used frequently throughout the paper.

Remark 3.4. The above definition is sometimes known as true complexity. It is known that a pattern (L_1, \dots, L_m) is complexity at most d if and only if $L_1^{d+1}, \dots, L_m^{d+1}$ are linearly independent as $(d+1)$ th-order tensors [18, 24].

Let (L_1, \dots, L_m) be any pattern such that no form is identically zero and no two forms are linearly dependent. It is known (for example, because Cauchy–Schwarz complexity is an upper bound for true complexity [22]) that (L_1, \dots, L_m) has complexity at most $m-2$.

It follows from the above discussion that the definition of complexity given in Definition 3.3 agrees with the definition of finite complexity given in Definition 2.6.

3.2. Nonclassical polynomials and homogeneity. The main technical result which is the basis for all tools in higher-order Fourier analysis is the inverse theorem for the Gowers uniformity norms (stated as Theorem 4.1 in this paper). This result states that a function which is not quasirandom with respect to the Gowers U^{d+1} -norm has some sort of structure. The specific structure involves the notion of nonclassical polynomials which we will need to work with heavily in this paper.

For ease of notation we write

$$(3.2) \quad \mathbb{U}_k := \frac{1}{p^k} \mathbb{Z}/\mathbb{Z} \subset \mathbb{R}/\mathbb{Z}$$

throughout the paper.

DEFINITION 3.5. Fix a prime p and a nonnegative integer $d \geq 0$. Let V be a finite-dimensional \mathbb{F}_p -vector space. A nonclassical polynomial of degree at most d is a map $P: V \rightarrow \mathbb{R}/\mathbb{Z}$ that satisfies

$$(D_{h_1} \cdots D_{h_{d+1}} P)(x) = 0$$

for all $h_1, \dots, h_{d+1}, x \in V$. The degree of P is the smallest $d > 0$ such that the above holds. The depth of P is the smallest $k \geq 0$ such that P takes values in a coset of \mathbb{U}_{k+1} .

See [34, Lemma 1.7] for some basic facts about nonclassical polynomials. We record one such fact here.

LEMMA 3.6 ([34, Lemma 1.7(iii)]). Fix a prime p and a finite-dimensional \mathbb{F}_p -vector space $V \simeq \mathbb{F}_p^n$. Then $P: V \rightarrow \mathbb{R}/\mathbb{Z}$ is a nonclassical polynomial of degree at most d if and only if it can be expressed in the form

$$P(x_1, \dots, x_n) = \alpha + \sum_{\substack{0 \leq i_1, \dots, i_n < p, j \geq 0: \\ 0 < i_1 + \dots + i_n \leq d - k(p-1)}} \frac{c_{i_1, \dots, i_n, k} |x_1|^{i_1} \cdots |x_n|^{i_n}}{p^{k+1}} \pmod{1}$$

for some $\alpha \in \mathbb{R}/\mathbb{Z}$ and coefficients $c_{i_1, \dots, i_n, k} \in \{0, \dots, p-1\}$ where $|\cdot|$ is the standard map $\mathbb{F}_p \rightarrow \{0, \dots, p-1\}$. Furthermore, this representation is unique.

As a corollary we see that in characteristic p , every nonclassical polynomial of degree at most d has depth at most $\lfloor (d-1)/(p-1) \rfloor$.

For technical reasons we will often need to work with homogeneous nonclassical polynomials. These are defined analogously to usual homogeneous polynomials and satisfy many of the expected properties.

DEFINITION 3.7. A homogeneous nonclassical polynomial is a nonclassical polynomial $P: V \rightarrow \mathbb{R}/\mathbb{Z}$ that also satisfies the following. For all $b \in \mathbb{F}_p$ there exists $\sigma_b^{(P)} \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ such that $P(bx) = \sigma_b^{(P)} P(x)$ for all $x \in V$.

As an example, consider the nonclassical cubic polynomial $P: \mathbb{F}_3 \rightarrow \mathbb{R}/\mathbb{Z}$ defined by $P(x) = |x|/9$. This is not homogeneous since $2P(1) = P(2)$, but $2P(2) \neq P(1)$. However, we can write $P = Q_3 + Q_2$ where $Q_3 = |x|/9 + |x|^2/3$ and $Q_2(x) = -|x|^2/3$. One can check that these are homogeneous nonclassical polynomials of degree 3 and 2, respectively.

LEMMA 3.8 ([24, Lemma 3.3]). Fix a prime p and integers $d > 0$ and $k \geq 0$ satisfying $k \leq \lfloor (d-1)/(p-1) \rfloor$. For each $b \in \mathbb{F}_p$ there exists $\sigma_b^{(d,k)} \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ such that $\sigma_b^{(P)} = \sigma_b^{(d,k)}$ for all homogeneous nonclassical polynomials P of degree d and depth k . Furthermore, for $b \neq 0$, the number $\sigma_b^{(d,k)}$ is uniquely determined by the following two properties:

- (i) $\sigma_b^{(d,k)} \equiv b^d \pmod{p}$,
- (ii) $(\sigma_b^{(d,k)})^{p-1} = 1$ (in $\mathbb{Z}/p^{k+1}\mathbb{Z}$).

THEOREM 3.9 ([24, Theorem 3.4]). *Let P be a nonclassical polynomial of degree d and depth k . Then P can be written as the sum of homogeneous nonclassical polynomials of degree at most d and depth at most k .*

3.3. Polynomial factors. The main objects we work with are polynomial factors, which are lists of polynomials which partition \mathbb{F}_p^n into pieces called atoms. There is a lot of data associated with a polynomial factor, so we introduce some notation which we use to keep track of this data.

DEFINITION 3.10. *Fix a prime p . Define*

$$D_p := \{(d, k) \in \mathbb{Z}_{>0} \times \mathbb{Z}_{\geq 0} : k \leq \lfloor (d-1)/(p-1) \rfloor\}$$

and

$$\mathcal{I}_p := \left\{ I \in \mathbb{Z}_{\geq 0}^{D_p} : \sum_{(d,k) \in D_p} I_{d,k} < \infty \right\}.$$

We call $I \in \mathcal{I}_p$ a parameter list. For $I \in \mathcal{I}_p$, we write $\|I\| := p^{\sum_{d,k} (k+1)I_{d,k}}$ and $\deg I$ for the largest d such that $I_{d,k} \neq 0$ for some k . We add and subtract parameter lists coordinatewise. For $I, I' \in \mathcal{I}_p$, we write $I \leq I'$ if $I_{d,k} \leq I'_{d,k}$ for all $(d, k) \in D_p$.

DEFINITION 3.11. *For p a prime and $I \in \mathcal{I}_p$, define the atom-indexing set of I to be*

$$(3.3) \quad A_I := \prod_{(d,k) \in D_p} \left(\frac{1}{p^{k+1}} \mathbb{Z}/\mathbb{Z} \right)^{I_{d,k}}.$$

Note that $|A_I| = \|I\|$.

For $I, I' \in \mathcal{I}_p$ with $I \leq I'$, write $\pi: A_{I'} \rightarrow A_I$ for the standard projection map defined by

$$(3.4) \quad \pi \left((a_{d,k}^i)_{\substack{(d,k) \in D_p \\ i \in [I'_{d,k}]}} \right) \mapsto (a_{d,k}^i)_{\substack{(d,k) \in D_p \\ i \in [I_{d,k}]}},$$

A_I is equipped with the following \mathbb{F}_p^\times -action:

$$(3.5) \quad c \cdot (a_{d,k}^i)_{\substack{(d,k) \in D_p \\ i \in [I_{d,k}]}} := \left(\sigma_c^{(d,k)} a_{d,k}^i \right)_{\substack{(d,k) \in D_p \\ i \in [I_{d,k}]}} ,$$

where $\sigma_c^{(d,k)}$ is defined in Lemma 3.8.

DEFINITION 3.12. *Fix a prime p . Let V be a finite-dimensional \mathbb{F}_p -vector space and let $I \in \mathcal{I}_p$ be a parameter list. A polynomial factor on V with parameters I , denoted \mathfrak{B} , is a collection*

$$(P_{d,k}^i)_{\substack{(d,k) \in D_p \\ i \in [I_{d,k}]}} ,$$

where $P_{d,k}^i$ is a homogeneous nonclassical polynomial of degree d and depth k . We also use \mathfrak{B} to denote the map $\mathfrak{B}: V \rightarrow A_I$ defined by evaluation of the polynomials. We also associate to \mathfrak{B} the partition of V given by the fibers of this map. The atoms of this partition are called the atoms of \mathfrak{B} . We write $\|\mathfrak{B}\| := \|I\|$ and $\deg \mathfrak{B} := \deg I$.

Note that if \mathfrak{B} is a polynomial factor on V with parameters I , then $\mathfrak{B}(cx) = c \cdot \mathfrak{B}(x)$ for all $c \in \mathbb{F}_p^\times$ and $x \in V$ where the \mathbb{F}_p^\times -action on A_I is defined in (3.5).

DEFINITION 3.13. Fix a prime p . Let V be a finite-dimensional \mathbb{F}_p -vector space and let $I, I' \in \mathcal{I}_p$ be two parameter lists. Let \mathfrak{B} and \mathfrak{B}' be two polynomial factors on V with parameters I and I' . We say that \mathfrak{B}' is a refinement of \mathfrak{B} if $I \leq I'$ and the lists of polynomials defining \mathfrak{B}' are extensions of the lists of polynomials defining \mathfrak{B} . We say that \mathfrak{B}' is a weak refinement of \mathfrak{B} if the partition of V associated to \mathfrak{B}' is a refinement of the partition associated to \mathfrak{B} .

Note that if \mathfrak{B}' is a refinement of \mathfrak{B} , then $\mathfrak{B} = \pi \circ \mathfrak{B}'$ where $\pi: A_{I'} \rightarrow A_I$ is the projection defined in (3.4).

We would like our polynomial factors to satisfy certain regularity conditions. For example, it will be convenient to know that all atoms are approximately the same size. All the regularity condition we need will be guaranteed if the factor is “high rank.”

DEFINITION 3.14. Fix a prime p and integer $d \geq 0$. Let V be a finite-dimensional \mathbb{F}_p -vector space. For a nonclassical polynomial $P: V \rightarrow \mathbb{R}/\mathbb{Z}$, define the d -rank of P , denoted $\text{rank}_d P$, to be the smallest integer r such that there exists non-classical polynomials $Q_1, \dots, Q_r: V \rightarrow \mathbb{R}/\mathbb{Z}$ of degree at most $d-1$ and a function $\Gamma: (\mathbb{R}/\mathbb{Z})^r \rightarrow \mathbb{R}/\mathbb{Z}$ such that $P(x) = \Gamma(Q_1(x), \dots, Q_r(x))$ for all $x \in V$.

For a polynomial factor \mathfrak{B} on V with parameters $I \in \mathcal{I}_p$, defined by a collection $(P_{d,k}^i)_{(d,k) \in D_p, i \in [I_{d,k}]}$ where $P_{d,k}^i$ is a homogeneous nonclassical polynomial of degree d and depth k , we define the rank of \mathfrak{B} , denoted $\text{rank } \mathfrak{B}$, to be

$$\min_{\lambda \in \prod_{(d,k) \in D_p} (\mathbb{Z}/p^{k+1}\mathbb{Z})^{I_{d,k}}} \text{rank}_{d'} \left(\sum_{(d,k) \in D_p} \sum_{i=1}^{I_{d,k}} \lambda_{d,k}^i P_{d,k}^i \right),$$

where

$$d' := \min_{(d,k) \in D_p, i \in [I_{d,k}]} \deg(\lambda_{d,k}^i P_{d,k}^i).$$

We need a couple of easy properties of rank, the first simply being that some high-rank factor exists.

LEMMA 3.15. Fix a prime p , a parameter list $I \in \mathcal{I}_p$, and a positive integer r . There exists a constant $n_{\text{high-rank}}(p, I, r)$ such that for every $n \geq n_{\text{high-rank}}(p, I, r)$, there exists a polynomial factor \mathfrak{B} on \mathbb{F}_p^n with parameters I and satisfying $\text{rank } \mathfrak{B} \geq r$.

Proof. For $(d, k) \in D_p$, write $d = (k + a)(p - 1) + b$ where $a \geq 0$ and $b \in \{1, \dots, p - 1\}$. By Lemma 3.6 there exists a nonclassical polynomial of degree d and depth k . For example, consider

$$\frac{|x_1|^{p-1} |x_2|^{p-1} \cdots |x_a|^b}{p^{k+1}} \pmod{1}.$$

By Theorem 3.9, we can decompose this nonclassical polynomial as the sum of homogeneous nonclassical polynomials of degree at most d and depth at most k . Let $P: \mathbb{F}_p^a \rightarrow \mathbb{U}_{k+1}$ be the homogeneous part of degree d and depth k .

Next define $Q: \mathbb{F}_p^{a \oplus N} \rightarrow \mathbb{U}_{k+1}$ by

$$Q(\mathbf{x}_1, \dots, \mathbf{x}_N) := P(\mathbf{x}_1) + \cdots + P(\mathbf{x}_N).$$

This is clearly a homogeneous nonclassical polynomial of degree d and depth k . We claim that for N large enough, we have $\text{rank}_d Q \geq r$. The proof of this fact uses several basic results from [34] that are not used elsewhere in this paper.

First, since P has degree exactly d , we have that $(D_{h_1} \cdots D_{h_d} P)(x)$ is a constant, independent of x , but is not identically zero. This implies that

$$c := \mathbb{E}_{h_1, \dots, h_d \in \mathbb{F}_p^a} e^{2\pi i (D_{h_1} \cdots D_{h_d} P)} < 1.$$

The quantity $-\log_p c$ is known as the *analytic rank* of P . Now a simple calculation shows that

$$\mathbb{E}_{h_1, \dots, h_d \in (\mathbb{F}_p^a)^{\oplus N}} e^{2\pi i (D_{h_1} \cdots D_{h_d} Q)} = c^N.$$

To conclude we use [34, Lemma 1.15(iii)], which implies that $\text{rank}_d Q \geq -c_d \log_p(c^N)$ for some constant $c_d > 0$ only depending on d . Since $c < 1$, taking N large enough gives $\text{rank}_d Q \geq r$, as desired.

Thus there exist homogeneous nonclassical polynomials $Q_{d,k}: V_{d,k} \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k that satisfy $\text{rank}_d Q_{d,k} \geq r$ for each $(d, k) \in D_p$. Define the vector space

$$V := \bigoplus_{(d,k) \in D_p} V_{d,k}^{\oplus I_{d,k}}.$$

Define the homogeneous nonclassical polynomials $\overline{Q}_{d,k}^i: V \rightarrow \mathbb{R}/\mathbb{Z}$ for each $(d, k) \in D_p$ and $i \in [I_{d,k}]$ such that $\overline{Q}_{d,k}^i$ is equal to $Q_{d,k}$ evaluated on the i th copy of $V_{d,k}$ and does not depend on the other coordinates. In particular, we define

$$\overline{Q}_{d,k}^i \left((x_{d',k'}^{i'})_{\substack{(d',k') \in D_p \\ i' \in [I_{d',k'}]}} \right) := Q_{d,k}(x_{d,k}^i).$$

These polynomials define a polynomial factor \mathfrak{B} on V with parameters I such that each of the homogeneous nonclassical polynomials defining \mathfrak{B} has rank at least r . Furthermore, since the polynomials defining \mathfrak{B} depend on disjoint sets of variables, it follows that all nontrivial linear combinations of the polynomials defining \mathfrak{B} also have high rank. Setting $n_{\text{high-rank}}(p, I, R) := \dim V$, we have constructed the desired polynomial factor on $\mathbb{F}_p^{n_{\text{high-rank}}(p, I, R)}$. To extend this construction to \mathbb{F}_p^n with $n \geq n_{\text{high-rank}}(p, I, r)$ one can simply add on extra variables that none of the polynomials depend on. \square

LEMMA 3.16 ([7, Lemma 2.13]). *Fix a prime p and positive integers d, r . Let V be a finite-dimensional \mathbb{F}_p -vector space and let $P: V \rightarrow \mathbb{R}/\mathbb{Z}$ be a nonclassical polynomial of degree d such that $\text{rank}_d(P) \geq r + p$. Let $U \leq V$ be a codimension-1 hyperplane. Then $\text{rank}_d(P|_U) \geq r$ unless $d = 1$ and $P|_U$ is identically zero.*

As a consequence, let \mathfrak{B} be a polynomial factor on V and let $P: V \rightarrow \mathbb{R}/\mathbb{Z}$ be a linear polynomial. Write \mathfrak{B}' for the common refinement of \mathfrak{B} and $\{P\}$. If $\text{rank } \mathfrak{B}' \geq r + p$, then $\text{rank } \mathfrak{B}|_U \geq r$ where U is the codimension-1 hyperplane where P vanishes.

3.4. Equidistribution and consistency sets. In this section we state a very useful result known as equidistribution and introduce several related notations. One consequence of equidistribution is the previously mentioned fact that high-rank polynomial factors have atoms of approximately equal sizes.

DEFINITION 3.17. Fix a prime p , integers $d > 0$ and $k \geq 0$ satisfying $k \leq \lfloor (d-1)/(p-1) \rfloor$, and a system $\mathbf{L} = (L_1, \dots, L_m)$ of m linear forms in ℓ variables. Define the (d, k) -consistency set of \mathbf{L} , denoted $\Phi_{d,k}(\mathbf{L})$, to be the subset of \mathbb{U}_{k+1}^m consisting of the tuples $\mathbf{a} = (a_1, \dots, a_m)$ such that there exists a finite-dimensional \mathbb{F}_p -vector space V , a homogeneous nonclassical polynomial $P: V \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k , and a tuple $\mathbf{x} \in V^\ell$ such that $a_i = P(L_i(\mathbf{x}))$ for all $i \in [m]$.

For a parameter list $I \in \mathcal{I}_p$, define the I -consistency set of \mathbf{L} to be the set of tuples $\mathbf{a} = (a_1, \dots, a_m) \in A_I^m$ such that for each $(d, k) \in D_p$ and $j \in [I_{d,k}]$ the tuple $((a_1)_{d,k}^j, \dots, (a_m)_{d,k}^j)$ lies in $\Phi_{d,k}(\mathbf{L})$.

LEMMA 3.18. Fix a prime p , integers $d > 0$ and $k \geq 0$ satisfying $k \leq \lfloor (d-1)/(p-1) \rfloor$, and a system $\mathbf{L} = (L_1, \dots, L_m)$ of m linear forms. The (d, k) -consistency set of \mathbf{L} is a subgroup of \mathbb{U}_{k+1}^m .

Proof. Suppose $\mathbf{a}, \mathbf{b} \in \Phi_{d,k}(\mathbf{L})$. We wish to show that $-\mathbf{a}$ and $\mathbf{a} + \mathbf{b}$ both lie in this set. By definition, there exist finite-dimensional \mathbb{F}_p -vector spaces V, W , homogeneous nonclassical polynomials $P: V \rightarrow \mathbb{U}_{k+1}$ and $Q: W \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k , and tuples $\mathbf{x} \in V^\ell$ and $\mathbf{y} \in W^\ell$ such that $a_i = P(L_i(\mathbf{x}))$ and $b_i = Q(L_i(\mathbf{y}))$ for all $i \in [m]$.

Note that $-\mathbf{a} \in \Phi_{d,k}(\mathbf{L})$ since $-P: V \rightarrow \mathbb{U}_{k+1}$ is a homogeneous nonclassical polynomial of degree d and depth k that satisfies $(-P)(L_i(\mathbf{x})) = -a_i$.

Now define $P \oplus Q: V \oplus W \rightarrow \mathbb{U}_{k+1}$ by $(P \oplus Q)(v \oplus w) := P(v) + Q(w)$. One can easily check that $P \oplus Q$ is a homogeneous nonclassical polynomial of degree d and depth k . Finally note that $(P \oplus Q)(L_i(\mathbf{x} \oplus \mathbf{y})) = a_i + b_i$, as desired.² \square

THEOREM 3.19 (equidistribution [24, Theorem 3.10]). Fix a prime p , a positive integer $d > 0$, and a parameter $\epsilon > 0$. There exists $r_{\text{equi}}(p, d, \epsilon)$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space and let \mathfrak{B} be a polynomial factor on V with parameters I such that $\deg \mathfrak{B} \leq d$ and $\text{rank}(\mathfrak{B}) \geq r_{\text{equi}}(p, d, \epsilon)$. Then for a system of linear forms $\mathbf{L} = (L_1, \dots, L_m)$ consisting of m forms in ℓ variables, and a tuple of atoms $\mathbf{a} = (a_1, \dots, a_m) \in \Phi_I(\mathbf{L})$,

$$\left| \Pr_{\mathbf{x} \in V^\ell} (\mathfrak{B}(L_i(\mathbf{x})) = a_i \text{ for all } i \in [m]) - \frac{1}{|\Phi_I(\mathbf{L})|} \right| \leq \epsilon.$$

Remark 3.20. To be completely correct, the statement given above follows by combining [24, Theorem 3.10] and [24, Corollary 2.13].

Note that the probability above is 0 if $\mathbf{a} \notin \Phi_I(\mathbf{L})$. We typically apply the above theorem with ϵ that decreases rapidly with $\|\mathbf{L}\|$, for example, taking $\epsilon = 1/(2\|\mathbf{L}\|^m)$ and using the fact that $|\Phi_I(\mathbf{L})| \leq \|\mathfrak{B}\|^m$, we see that in this case the probability above is at least $1/(2|\Phi_I(\mathbf{L})|)$.

Consistency sets are often hard to compute exactly. The next two lemmas give exact relations on the sizes of consistency sets in two special cases that occur in this paper.

DEFINITION 3.21. Fix a prime p . A system \mathbf{L} of m linear forms in ℓ variables over \mathbb{F}_p is full dimensional if $|\Phi_{d,k}(\mathbf{L})| = |\Phi_{d,k}(\mathbf{L}^\ell)|$ for all $(d, k) \in D_p$ (recall the system \mathbf{L}^ℓ defined in Definition 2.7 defines an ℓ -dimensional subspace).

²To be completely correct, we also need to show that $\Phi_{d,k}(\mathbf{L})$ is nonempty, which follows from, for example, Lemma 3.6 and Theorem 3.9, which together show the existence of a homogeneous non-classical polynomial of degree d and depth k for every $(d, k) \in D_p$.

LEMMA 3.22. Fix a prime p and a positive integer ℓ . Let $J \subseteq \mathbb{F}_p^\ell$ be a set that contains at least one vector in each direction (i.e., for each $\mathbf{i} \in \mathbb{F}_p^\ell$, there exists $\mathbf{j} \in J$ and $b \in \mathbb{F}_p^\times$ such that $\mathbf{i} = b\mathbf{j}$). Consider the system $\mathbf{L}_J := (L_{\mathbf{i}}^\ell)_{\mathbf{i} \in J}$ of $|J|$ linear forms in ℓ variables (recall the linear form $L_{\mathbf{i}}^\ell$ defined in Definition 2.7). Then \mathbf{L}_J is full dimensional.

As a special case of this result we see that the system $\overline{\mathbf{L}}^\ell$, defined in Definition 2.7, is full dimensional.

Proof. Note that the system \mathbf{L}_J is a subsystem of \mathbf{L}^ℓ . This immediately implies that $|\Phi_{d,k}(\mathbf{L}_J)| \leq |\Phi_{d,k}(\mathbf{L}^\ell)|$, since if $(a_{\mathbf{i}})_{\mathbf{i} \in \mathbb{F}_p^\ell} \in \Phi_{d,k}(\mathbf{L}^\ell)$, then $(a_{\mathbf{i}})_{\mathbf{i} \in J} \in \Phi_{d,k}(\mathbf{L}_J)$.

To go the other direction, we use the homogeneity of our polynomials. Suppose $(a_{\mathbf{j}})_{\mathbf{j} \in J} \in \Phi_{d,k}(\mathbf{L}_J)$. Thus there exists a finite-dimensional \mathbb{F}_p -vector space V and a homogeneous nonclassical polynomial $P: V \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k , and a vector $\mathbf{x} \in V^\ell$ such that $P(L_{\mathbf{j}}^\ell(\mathbf{x})) = a_{\mathbf{j}}$ for all $\mathbf{j} \in J$. Now by assumption, for $\mathbf{i} \in \mathbb{F}_p^\ell$, there exists $\mathbf{j} \in J$ and $b \in \mathbb{F}_p$ such that $\mathbf{i} = b\mathbf{j}$. Define $a_{\mathbf{i}} := \sigma_b^{(d,k)} a_{\mathbf{j}}$ (recall the definition of σ_b from Lemma 3.8). We claim that $(a_{\mathbf{i}})_{\mathbf{i} \in \mathbb{F}_p^\ell} \in \Phi_{d,k}(\mathbf{L}^\ell)$. This is true simply because

$$P(L_{\mathbf{i}}^\ell(\mathbf{x})) = P(bL_{\mathbf{j}}^\ell(\mathbf{x})) = \sigma_b^{(d,k)} P(L_{\mathbf{j}}^\ell(\mathbf{x})) = \sigma_b^{(d,k)} a_{\mathbf{j}} = a_{\mathbf{i}},$$

where $\mathbf{j} \in J$ and $b \in \mathbb{F}_p$ are defined as above. Thus $|\Phi_{d,k}(\mathbf{L}_J)| \geq |\Phi_{d,k}(\mathbf{L}^\ell)|$, as desired. \square

Several times in the paper, an application of the Cauchy–Schwarz inequality will lead us to consider counts of triples of patterns $\mathbf{L}, \mathbf{L}', \mathbf{L}''$ of the following form.

LEMMA 3.23. Fix a prime p . Let \mathbf{L} be a system of m linear forms in ℓ variables over \mathbb{F}_p . Say \mathbf{L} is defined by M , an $m \times \ell$ matrix. (By this we mean that $L_i(x_1, \dots, x_\ell) = M_{i,1}x_1 + \dots + M_{i,\ell}x_\ell$ for all $i \in [m]$.) Let \mathbf{L}' be a system of $m(n+1)$ linear forms in $\ell + \ell'$ variables, defined by a matrix of the form

$$\left(\begin{array}{c|c} M & 0 \\ \hline c_1 M & \\ \vdots & \\ c_n M & \end{array} \middle| \begin{array}{c} N \end{array} \right),$$

where $c_1, \dots, c_n \in \mathbb{F}_p$ and N is an $mn \times \ell'$ matrix. Let \mathbf{L}'' be the system of $m(2n+1)$ linear forms in $\ell + 2\ell'$ variables, defined by the matrix

$$\left(\begin{array}{c|c|c} M & 0 & 0 \\ \hline c_1 M & & \\ \vdots & N & 0 \\ \hline c_n M & & \\ \hline c_1 M & 0 & N \\ \vdots & & \\ c_n M & & \end{array} \right).$$

Then for all $(d, k) \in D_p$, we have

$$|\Phi_{d,k}(\mathbf{L})| \cdot |\Phi_{d,k}(\mathbf{L}'')| = |\Phi_{d,k}(\mathbf{L}')|^2.$$

Proof. We construct injections between $\Phi_{d,k}(\mathbf{L}) \times \Phi_{d,k}(\mathbf{L}'')$ and $\Phi_{d,k}(\mathbf{L}') \times \Phi_{d,k}(\mathbf{L}')$ in both directions. Write $\sigma_i := \sigma_{c_i}^{d,k}$ for $i \in [n]$ for the rest of the proof (see Lemma 3.8 for the definition).

Consider $\mathbf{a} = (a_1, \dots, a_m) \in \Phi_{d,k}(\mathbf{L})$ and $\mathbf{b} = (b_1, \dots, b_{m(2n+1)}) \in \Phi_{d,k}(\mathbf{L}'')$. By definition, there exists a finite-dimensional \mathbb{F}_p -vector space V , a homogeneous nonclassical polynomial $P: V \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k , and a vector $\mathbf{x} \in V^\ell$ such that $P(L_i(\mathbf{x})) = a_i$ for all $i \in [m]$. Also by definition, there exists a finite-dimensional \mathbb{F}_p -vector space W , a homogeneous nonclassical polynomial $Q: W \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k , and a vector $(\mathbf{x}', \mathbf{y}, \mathbf{y}') \in W^\ell \times W^{\ell'} \times W^{\ell'}$ such that $Q(L'_i(\mathbf{x}', \mathbf{y}, \mathbf{y}')) = b_i$ for all $i \in [m(2n+1)]$.

Now we map (\mathbf{a}, \mathbf{b}) to the pair $(\mathbf{a}', \mathbf{b}')$ where $b'_i = b_i$ for $i \in [m(n+1)]$ and $a'_i = a_i + b_i$ for $i \in [m]$ and $a'_{tm+i} = \sigma_t a_i + b_{m(n+t)+i}$ for $i \in [m]$ and $t \in [n]$. We can easily check that no two pairs (\mathbf{a}, \mathbf{b}) map to the same pair $(\mathbf{a}', \mathbf{b}')$. All that remains is to check that $\mathbf{a}', \mathbf{b}' \in \Phi_{d,k}(\mathbf{L}')$.

Define $P \oplus Q: V \oplus W \rightarrow \mathbb{U}_{k+1}$ by $(P \oplus Q)(x \oplus y) := P(x) + Q(y)$. This is clearly a homogeneous nonclassical polynomial of degree d and depth k . Note that $\mathbf{z} := (\mathbf{x} \oplus \mathbf{x}', \mathbf{0} \oplus \mathbf{y}') \in (V \oplus W)^{\ell+\ell'}$ satisfies $(P \oplus Q)(L'_i(\mathbf{z})) = a'_i$ for all $i \in [m(n+1)]$. Similarly note that $\mathbf{z}' := (\mathbf{0} \oplus \mathbf{x}', \mathbf{0} \oplus \mathbf{y}) \in (V \oplus W)^{\ell+\ell'}$ satisfies $(P \oplus Q)(L'_i(\mathbf{z}')) = b'_i$ for all $i \in [m(n+1)]$. This demonstrates the first injection.

Now consider $\mathbf{a} = (a_1, \dots, a_{m(n+1)}) \in \Phi_{d,k}(\mathbf{L}')$ and $\mathbf{b} = (b_1, \dots, b_{m(n+1)}) \in \Phi_{d,k}(\mathbf{L}')$. By definition, there exists a finite-dimensional \mathbb{F}_p -vector space V , a homogeneous nonclassical polynomial $P: V \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k , and a vector $(\mathbf{x}, \mathbf{y}) \in V^\ell \times V^{\ell'}$ such that $P(L'_i(\mathbf{x}, \mathbf{y})) = a_i$ for $i \in [m(n+1)]$. Also by definition, there exists a finite-dimensional \mathbb{F}_p -vector space W , a homogeneous nonclassical polynomial $Q: W \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k , and a vector $(\mathbf{x}', \mathbf{y}') \in W^\ell \times W^{\ell'}$ such that $Q(L'_i(\mathbf{x}', \mathbf{y}')) = b_i$ for $i \in [m(n+1)]$.

We map (\mathbf{a}, \mathbf{b}) to the pair $(\mathbf{a}', \mathbf{b}')$ where $a'_i = a_i$ for $i \in [m]$ and $b'_i = a_i + b_i$ for $i \in [m]$ and $b'_{m+i} = a_{m+i}$ for $i \in [mn]$ and $b'_{m(n+1)+i} = b_{m+i}$ for $i \in [mn]$. We can easily check that no two pairs (\mathbf{a}, \mathbf{b}) map to the same pair $(\mathbf{a}', \mathbf{b}')$. All that remains is to check that $\mathbf{a}' \in \Phi_{d,k}$ and $\mathbf{b}' \in \Phi_{d,k}(\mathbf{L}'')$.

As above, define the homogeneous nonclassical polynomial $P \oplus Q: V \oplus W \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k by $(P \oplus Q)(x \oplus y) := P(x) + Q(y)$. Note that $\mathbf{z} := \mathbf{x} \oplus \mathbf{0} \in (V \oplus W)^\ell$ satisfies $(P \oplus Q)(L_i(\mathbf{z})) = a'_i$ for $i \in [m]$ and $\mathbf{z}' := (\mathbf{x} \oplus \mathbf{x}', \mathbf{y} \oplus \mathbf{0}, \mathbf{0} \oplus \mathbf{y}') \in (V \oplus W)^\ell \times (V \oplus W)^{\ell'} \times (V \oplus W)^{\ell'}$ satisfies $(P \oplus Q)(L'_i(\mathbf{z}')) = b'_i$ for $i \in [m(2n+1)]$, as desired. \square

3.5. Subatom selection functions. A situation that often occurs is the following. We have a polynomial factor \mathfrak{B} with parameters I and a refinement \mathfrak{B}' with parameters I' . We use the word atom to refer to the atoms of the partition induced by \mathfrak{B} ; these atoms are indexed by A_I . We use the word subatom to refer to the atoms of the partition induced by \mathfrak{B}' ; these atoms are indexed by $A_{I'}$. The projection map $\pi: A_{I'} \rightarrow A_I$, defined in (3.4), maps a subatom to the atom that it is contained in.

We wish to designate one subatom inside each atom as special. This choice is given by a map $s: A_I \rightarrow A_{I'}$ that is a right inverse for π . In this paper we define a certain class of these maps that we call subatom selection functions that have several desirable properties.

First we define certain polynomials $P_{d,k}: \mathbb{F}_p \rightarrow \mathbb{U}_{k+1}$ for each $(d, k) \in D_p$. Recall from Definition 3.10 that $(d, k) \in D_p$ means that $d \geq k(p-1) + 1$. First, for $k(p-1) + 1 \leq d \leq (k+1)(p-1)$, let $P_{d,k}$ be an arbitrary homogeneous nonclassical polynomial $\mathbb{F}_p \rightarrow \mathbb{U}_{k+1}$ of degree d and depth k in one variable. (The existence of such

a polynomial follows from Lemma 3.6 and Theorem 3.9.) Then for the remaining cases, i.e., $d > (k+1)(p-1)$, simply define $P_{d,k} := P_{d',k}$ where $k(p-1)+1 \leq d' \leq (k+1)(p-1)$ satisfies $d' \equiv d \pmod{p-1}$.

DEFINITION 3.24. Fix a prime p and parameter lists $I, I' \in \mathcal{I}_p$ satisfying $I \leq I'$. Let $c_{d,k}^{i,j} \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ be arbitrary elements for $(d,k) \in D_p$ and $i \in [I_{1,0}]$ and $I_{d,k} < j \leq I'_{d,k}$. A subatom selection function is a map of the form $s_c: A_I \rightarrow A_{I'}$, defined by

$$[s_c(a)]_{d,k}^i = \begin{cases} a_{d,k}^i & \text{if } i \leq I_{d,k}, \\ \sum_{j=1}^{I_{1,0}} c_{d,k}^{j,i} P_{d,k}(|a_{1,0}^j|) & \text{otherwise,} \end{cases}$$

where the maps $P_{d,k}: \mathbb{F}_p \rightarrow \mathbb{U}_{k+1}$ were defined in the preceding paragraph and $|\cdot|$ is the standard map $\mathbb{U}_1 \rightarrow \mathbb{F}_p$.

LEMMA 3.25. Fix a prime p , parameter lists $I, I' \in \mathcal{I}_p$ satisfying $I \leq I'$, and a subatom selection function $s_c: A_I \rightarrow A_{I'}$. The following hold:

- (i) $\pi \circ s_c = \text{Id}$ (where $\pi: A_{I'} \rightarrow A_I$ is defined in (3.4));
- (ii) for $a \in A_I$ and $b \in \mathbb{F}_p^\times$, we have $b \cdot s_c(a) = s_c(b \cdot a)$ (where the action of \mathbb{F}_p^\times on A_I and $A_{I'}$ is defined in (3.5));
- (iii) for every system \mathbf{L} of m linear forms and every consistent tuple of atoms $(a_1, \dots, a_m) \in \Phi_I(\mathbf{L})$, we have

$$(s_c(a_1), \dots, s_c(a_m)) \in \Phi_{I'}(\mathbf{L})$$

(see Definition 3.17 for the definition of the consistency sets $\Phi_I(\mathbf{L})$ and $\Phi_{I'}(\mathbf{L})$).

Proof. Property (i) is immediate.

For property (ii), by definition, we have

$$[b \cdot s_c(a)]_{d,k}^i = \sigma_b^{(d,k)} [s_c(a)]_{d,k}^i,$$

where $\sigma_b^{(d,k)}$ is defined in Lemma 3.8. Now $s_c(b \cdot a)_{d,k}^i = \sigma_b^{(d,k)} a_{d,k}^i$ if $i \leq I_{d,k}$, so we are done in this case.

Assume otherwise. Define d' such that $d' \equiv d \pmod{p-1}$ and $k(p-1)+1 \leq d' \leq (k+1)(p-1)$. Remember that $P_{d,k} = P_{d',k}$ is a homogeneous nonclassical polynomial of degree d' and depth k . Also note that $\sigma_b^{(1,0)} = b \in \mathbb{Z}/p\mathbb{Z}$. Then we have

$$[s_c(b \cdot a)]_{d,k}^i = \sum_{j=1}^{I_{1,0}} c_{d,k}^{j,i} P_{d,k}(b|a_{1,0}^j|) = \sigma_b^{(d',k)} \sum_{j=1}^{I_{1,0}} c_{d,k}^{j,i} P_{d,k}(|a_{1,0}^j|).$$

To complete the proof of (ii) we need to show that $\sigma_b^{d,k} = \sigma_b^{d',k}$ whenever $d \equiv d' \pmod{p-1}$. This follows from Lemma 3.8, which implies that $\sigma_b^{d,k}$ is uniquely determined by the facts that $\sigma_b^{d,k} \equiv b^d \pmod{p}$ and $(\sigma_b^{(d,k)})^{p-1} = 1$ in $\mathbb{Z}/p^{k+1}\mathbb{Z}$. The first property does not change when d changes by a multiple of $p-1$ (by Fermat's little theorem) and the second property does not depend on d at all. Thus we conclude the desired result.

Now we prove (iii). We know, by (3.1), that the consistency set $\Phi_{d,k}(\mathbf{L})$ is a subgroup of \mathbb{U}_{k+1}^m . Thus it suffices to prove that for $(a_1, \dots, a_m) \in \Phi_{1,0}(\mathbf{L}) \subseteq \mathbb{U}_1^m$ we have $(P_{d,k}(|a_1|), \dots, P_{d,k}(|a_m|)) \in \Phi_{d,k}(\mathbf{L})$. Given that $(a_1, \dots, a_m) \in \Phi_{1,0}$ we know that

there exists a finite-dimensional \mathbb{F}_p -vector space V , a linear function $P: V \rightarrow \mathbb{U}_1$, and vectors $\mathbf{x} = (x_1, \dots, x_\ell) \in V^\ell$ such that $P(L_i(\mathbf{x})) = a_i$ for $i \in [m]$. Since P and L_i are linear, they commute, and thus $L_i(\mathbf{y}) = |a_i|$ for all $i \in [m]$ where $\mathbf{y} \in \mathbb{F}_p^\ell$ is defined by $y_i = |P(x_i)|$. Finally, since $P_{d,k}$ is a homogeneous nonclassical polynomial of degree d' and depth k , we have $(P_{d,k}(|a_1|), \dots, P_{d,k}(|a_m|)) = (P_{d,k}(L_1(\mathbf{y})), \dots, P_{d,k}(L_m(\mathbf{y}))) \in \Phi_{d',k}(\mathbf{L})$.

To complete the proof, we show that $\Phi_{d',k}(\mathbf{L}) \subseteq \Phi_{d,k}(\mathbf{L})$. Let $Q: \mathbb{F}_p^n \rightarrow \mathbb{U}_{k+1}$ be a homogeneous nonclassical polynomial of degree d and depth k . (This exists as long as $n \geq \lceil (d-1)/(p-1) \rceil - k$.) Then consider the map $P_{d,k} \oplus Q: \mathbb{F}_p \oplus \mathbb{F}_p^n \rightarrow \mathbb{U}_{k+1}$ defined as usual by $(P_{d,k} \oplus Q)(x \oplus y) := P_{d,k}(x) + Q(y)$. This is clearly a nonclassical polynomial of degree d and depth k . Furthermore,

$$(P_{d,k} \oplus Q)(b(x \oplus y)) = \sigma_b^{(d',k)} P_{d,k}(x) + \sigma_b^{(d,k)} Q(y) = \sigma_b^{(d,k)} (P_{d,k} \oplus Q)(x \oplus y)$$

since $d' \equiv d \pmod{p-1}$. Considering $\mathbf{y} \oplus \mathbf{0} \in (\mathbb{F}_p \oplus \mathbb{F}_p^n)^\ell$ shows that $(P_{d,k}(|a_1|), \dots, P_{d,k}(|a_m|)) \in \Phi_{d,k}(\mathbf{L})$, as desired. \square

4. Arithmetic regularity and subatom selection. In this section we prove the subatom selection result, Theorem 4.6, that will be one of the main technical tools used in the proof of our main theorem. To prove this result we first prove a strong arithmetic regularity lemma, Lemma 4.5, via a chain of regularity lemmas of varying strengths. From the strong regularity lemma we deduce the subatom selection result via the probabilistic method.

The first two regularity lemmas, Lemmas 4.3 and 4.4, are very similar to results in [8, 7], differing only in some technical details. The main innovation in this section is that Lemma 4.5 is much stronger than the corresponding result in [8, 7]. To accomplish this, we iterate Lemma 4.4 with the complexity parameter (i.e., degree of the nonclassical polynomials) increasing at each step of the iteration. To our knowledge, this idea has not appeared previously in the literature and it is crucial to handle patterns of unbounded complexity.

Notation and conventions. Recall that a polynomial factor \mathfrak{B} on a vector space V with parameters I gives rise to a partition (or σ -algebra) on V whose atoms are the fibers of the map $\mathfrak{B}: V \rightarrow A_I$. For a function $f: V \rightarrow \mathbb{C}$, we write $\mathbb{E}[f|\mathfrak{B}]: V \rightarrow \mathbb{C}$ for the projection of f onto the σ -algebra generated by \mathfrak{B} . Concretely, $\mathbb{E}[f|\mathfrak{B}](x)$ is defined to be the average of f over the atom of \mathfrak{B} which contains x .

In this section we have to deal with many growth functions. Without loss of generality we always assume that these growth functions are monotone in all their parameters.

4.1. Arithmetic regularity lemmas. Given functions $f^{(1)}, \dots, f^{(R)}: V \rightarrow [0, 1]$ and a polynomial factor \mathfrak{B} , we call the quantity

$$\sum_{\ell=1}^R \left| \mathbb{E}[f^{(\ell)}|\mathfrak{B}] \right|_2^2$$

the energy of \mathfrak{B} . The main technique we use in the proofs of regularity lemmas is an energy increment technique; using the fact that the energy is bounded between 0 and R , we show that some iteration must halt after a bounded number of steps.

By applying this iteration to the inverse theorem, we deduce a weak regularity lemma, Lemma 4.3, which produces a decomposition $f = f_{\text{str}} + f_{\text{psr}}$ where f_{str} is structured in the sense that $f_{\text{str}} = \mathbb{E}[f|\mathfrak{B}]$ for some polynomial factor \mathfrak{B} and f_{psr} is

U^{d+1} -pseudorandom. Furthermore, it will be important that we can make the rank of \mathfrak{B} arbitrarily large in terms of its size.

Iterating again gives a regularity lemma, Lemma 4.4, which produces a decomposition $f = f_{\text{str}} + f_{\text{psr}} + f_{\text{sml}}$ where f_{str} is structured, f_{psr} is U^{d+1} -pseudorandom, and f_{sml} is small in L^2 -norm. The main benefit of this result is that the amount of pseudorandomness can be arbitrarily small in terms of the size of \mathfrak{B} .

Iterating a third time gives a strong regularity lemma, Lemma 4.5 (sometimes called a “two-level decomposition theorem” or a “super decomposition theorem”), which produces a pair of polynomial factors \mathfrak{B} refined by \mathfrak{B}' and a decomposition $f = f_{\text{str}} + f_{\text{psr}} + f_{\text{sml}}$ where $f_{\text{str}} = \mathbb{E}[f|\mathfrak{B}']$. This result has two main strengths. First, $\|f_{\text{sml}}\|_2$ can be made arbitrarily small in terms of the size of \mathfrak{B} . And second, the main innovation of this section, f_{psr} is U^{d+1} -pseudorandom where d can be made arbitrarily larger in terms of the size of \mathfrak{B} . To accomplish the second property, in the proof we simply iterate Lemma 4.4 with the degree of the nonclassical polynomials involved increasing at each step of the iteration.

Other than this innovation, most of the proofs in this section are fairly standard in this field. We include their proofs for completeness and because the exact technical details we require differ slightly from other results that have appeared previously.

THEOREM 4.1 (inverse theorem [34, Theorem 1.10]). *Fix a prime p , a positive integer d , and a parameter $\delta > 0$. There exists $\epsilon_{\text{inv}}(p, d, \delta) > 0$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space. Given a function $f: V \rightarrow \mathbb{C}$ satisfying $\|f\|_\infty \leq 1$ and $\|f\|_{U^{d+1}} > \delta$, there exists a nonclassical polynomial $P: V \rightarrow \mathbb{R}/\mathbb{Z}$ of degree at most d such that*

$$\left| \mathbb{E}_{x \in V} f(x) e^{-2\pi i P(x)} \right| \geq \epsilon_{\text{inv}}(p, d, \delta).$$

The next lemma is important for making factors high rank and its second claim is critical in proving the stronger regularity lemma where we need to produce a refinement (instead of a weak refinement).

LEMMA 4.2 (making factors high rank [24]; cf. [7, Theorem 2.19]). *Fix a prime p , positive integers d, C_0 , and a nondecreasing function $r: \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$. There exist constants $C_{\text{rank}}(p, d, C_0, r)$ and $r_{\text{rank}}(p, d, C_0, r)$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space. Suppose that \mathfrak{B} and \mathfrak{B}' are polynomial factors on V with degree at most d such that \mathfrak{B}' refines \mathfrak{B} and $\|\mathfrak{B}'\| \leq C_0$ and*

$$\text{rank } \mathfrak{B} \geq r_{\text{rank}}(p, d, C_0, r).$$

Then there is a polynomial factor \mathfrak{B}'' on V that weakly refines \mathfrak{B}' , refines \mathfrak{B} , and satisfies $\|\mathfrak{B}''\| \leq C_{\text{rank}}(p, d, C_0, r)$ and $\deg \mathfrak{B}'' \leq d$ and $\text{rank } \mathfrak{B}'' \geq r(\|\mathfrak{B}''\|)$.

LEMMA 4.3 (weak arithmetic regularity). *Fix a prime p , positive integers d, R, C_0 , a parameter $\eta > 0$, and a nondecreasing function $r: \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$. There exist constants $C_{\text{reg}}(p, d, R, C_0, \eta, r)$ and $r_{\text{reg}}(p, d, R, C_0, \eta, r)$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space and let \mathfrak{B}_0 be a polynomial factor on V satisfying $\|\mathfrak{B}_0\| \leq C_0$ and $\deg \mathfrak{B}_0 \leq d$ and $\text{rank } \mathfrak{B}_0 \geq r_{\text{reg}}(p, d, R, C_0, \eta, r)$. Given functions $f^{(1)}, \dots, f^{(R)}: V \rightarrow [0, 1]$, there exists a polynomial factor \mathfrak{B} on V that refines \mathfrak{B}_0 with the following properties. There exists a decomposition*

$$f^{(\ell)} = f_{\text{str}}^{(\ell)} + f_{\text{psr}}^{(\ell)}$$

for each $\ell \in [R]$ such that

- (i) $f_{\text{str}}^{(\ell)} = \mathbb{E}[f^{(\ell)}|\mathfrak{B}]$ for each $\ell \in [R]$;
- (ii) $\|f_{\text{psr}}^{(\ell)}\|_{U^{d+1}} < \eta$ for each $\ell \in [R]$;
- (iii) $f_{\text{str}}^{(\ell)}$ has range $[0, 1]$ and $f_{\text{psr}}^{(\ell)}$ has range $[-1, 1]$ for each $\ell \in [R]$;
- (iv) $\text{rank } \mathfrak{B} \geq r(\|\mathfrak{B}\|)$;
- (v) $\|\mathfrak{B}\| \leq C_{\text{reg}}(p, d, R, C_0, \eta, r)$ and $\deg \mathfrak{B} \leq d$.

Proof. Set $M := \lceil R\epsilon_{\text{inv}}(p, d, \eta)^{-2} \rceil$. Define nondecreasing functions $r_i: \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ for each $i = 0, \dots, M$ such that $r_0 = r$ and $r_{i+1}(N) \geq r_{\text{rank}}(p, d, p^{d^3}N, r_i)$ and such that $r_{i+1}(N) \geq r_i(N)$ for all $i = 0, \dots, M-1$ and all $N \in \mathbb{Z}_{>0}$. Define $r_{\text{reg}}(p, d, R, C_0, \eta, r) := r_M(C_0)$.

We construct a list of polynomial factors $\mathfrak{B}_0, \mathfrak{B}_1, \dots, \mathfrak{B}_m$ on V such that

- \mathfrak{B}_i refines \mathfrak{B}_{i-1} for $i = 1, \dots, m$;
- $\text{rank } \mathfrak{B}_i \geq r_{M-i}(\|\mathfrak{B}_i\|)$ for $i = 0, \dots, m$;
- $\|\mathfrak{B}_i\| \leq C_{\text{rank}}(p, d, p^{d^3}\|\mathfrak{B}_{i-1}\|, r_{M-i})$ and $\deg \mathfrak{B}_i \leq d$ for $i = 1, \dots, m$.

Suppose we have constructed polynomial factors $\mathfrak{B}_0, \dots, \mathfrak{B}_i$ with the above properties. If $\|f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_{U^{d+1}} < \eta$ for each $\ell \in [R]$ we halt the iteration. Otherwise there is some $\ell \in [R]$ such that, writing $g := f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]$, we have

$$\|g\|_{U^{d+1}} \geq \eta.$$

By Theorem 4.1, there exists a nonclassical polynomial $P: V \rightarrow \mathbb{R}/\mathbb{Z}$ of degree at most d such that

$$\left| \mathbb{E}_{x \in V} g(x) e^{-2\pi i P(x)} \right| \geq \epsilon_{\text{inv}}(p, d, \eta).$$

By Theorem 3.9, we can write $P = P_1 + \dots + P_C$ as the sum of homogeneous nonclassical polynomials. There are at most $\sum_{i=1}^d 1 + \lfloor (i-1)/(p-1) \rfloor \leq d^2$ terms in this sum. Let \mathfrak{B}'_i be the polynomial factor defined by the polynomials defining \mathfrak{B} as well as the polynomials P_1, \dots, P_C . Note that $\|\mathfrak{B}'_i\| \leq p^{d^3}\|\mathfrak{B}_i\|$. Finally let \mathfrak{B}_{i+1} be the polynomial factor produced by applying Lemma 4.2 to \mathfrak{B}_i and \mathfrak{B}'_i with parameters p, d, r_{M-i-1} . In particular \mathfrak{B}'_i refines \mathfrak{B}_i and

$$\begin{aligned} \text{rank } \mathfrak{B}_i &\geq r_{M-i}(\|\mathfrak{B}_i\|) \\ &\geq r_{\text{rank}}(p, d, p^{d^3}\|\mathfrak{B}_i\|, r_{M-i-1}) \\ &\geq r_{\text{rank}}(p, d, \|\mathfrak{B}'_i\|, r_{M-i-1}), \end{aligned}$$

so the hypotheses of Lemma 4.2 are satisfied. Thus we have defined \mathfrak{B}_{i+1} with all the desired properties.

Next we claim that this iteration must stop after at most M steps. We claim that

$$\sum_{\ell=1}^R \left\| \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i] \right\|_2^2$$

increases by at least $\epsilon_{\text{inv}}(p, d, \eta)^2$ each time i increases. Since this sum is clearly bounded between 0 and R , it suffices to prove this claim.

First note that by the Cauchy–Schwarz inequality, $\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i+1}]\|_2^2 \geq \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2$ holds for all ℓ . Now pick $\ell \in [R]$ as in the i th iteration, define $g := f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]$, and

let P be the nonclassical polynomial defined in the i th iteration. Note in particular that $e^{-2\pi i P(x)}$ is in the σ -algebra defined by \mathfrak{B}'_i . Then we compute

$$\begin{aligned} \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i+1}]\|_2^2 - \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 &\geq \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}'_i]\|_2^2 - \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 \\ &= \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}'_i] - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 \\ &= \|\mathbb{E}[g|\mathfrak{B}'_i]\|_2^2 \\ &\geq \langle \mathbb{E}[g|\mathfrak{B}'_i], e^{2\pi i P} \rangle^2 \\ &= \langle g, e^{2\pi i P} \rangle^2 \\ &\geq \epsilon_{\text{inv}}(p, d, \eta)^2. \end{aligned}$$

Thus we have produced \mathfrak{B}_m with $m \leq M$ such that \mathfrak{B}_m refines \mathfrak{B}_0 and $\text{rank } \mathfrak{B}_m \geq r_{M-m}(\|\mathfrak{B}_m\|) \geq r(\|\mathfrak{B}_m\|)$ and $\|f^{(\ell)} - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_m]\|_{U^{d+1}} < \eta$ for each $\ell \in [R]$. Defining $f_{\text{str}}^{(\ell)} := \mathbb{E}[f^{(\ell)}|\mathfrak{B}_m]$ and $f_{\text{psr}}^{(\ell)} := f^{(\ell)} - f_{\text{str}}^{(\ell)}$, we immediately see that conclusions (i), (ii), (iii), and (iv) hold. Conclusion (v) holds by defining $C_{\text{reg}'}(p, d, R, C_0, \eta, r)$ to be the M -fold iteration of the function $N \mapsto C_{\text{rank}}(p, d, p^{d^3}N, r_M)$ applied to C_0 . \square

LEMMA 4.4 (arithmetic regularity). *Fix a prime p , positive integers d, R, C_0 , a parameter $\theta > 0$, a nonincreasing function $\eta: \mathbb{Z}_{>0} \rightarrow (0, 1)$, and a nondecreasing function $r: \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$. There exist constants $C_{\text{reg}''}(p, d, R, C_0, \theta, \eta, r)$ and $r_{\text{reg}''}(p, d, R, C_0, \theta, \eta, r)$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space and let \mathfrak{B}_0 be a polynomial factor on V satisfying $\|\mathfrak{B}_0\| \leq C_0$ and $\deg \mathfrak{B}_0 \leq d$ and $\text{rank } \mathfrak{B}_0 \geq r_{\text{reg}''}(p, d, R, C_0, \theta, \eta, r)$. Given functions $f^{(1)}, \dots, f^{(R)}: V \rightarrow [0, 1]$, there exists a polynomial factor \mathfrak{B} on V that refines \mathfrak{B}_0 with the following properties. There exists a decomposition*

$$f^{(\ell)} = f_{\text{str}}^{(\ell)} + f_{\text{psr}}^{(\ell)} + f_{\text{sml}}^{(\ell)}$$

for each $\ell \in [R]$ such that

- (i) $f_{\text{str}}^{(\ell)} = \mathbb{E}[f^{(\ell)}|\mathfrak{B}]$ for each $\ell \in [R]$;
- (ii) $\|f_{\text{psr}}^{(\ell)}\|_{U^{d+1}} < \eta(\|\mathfrak{B}\|)$ for each $\ell \in [R]$;
- (iii) $f_{\text{str}}^{(\ell)}$ and $f_{\text{str}}^{(\ell)} + f_{\text{sml}}^{(\ell)}$ have range $[0, 1]$ and $f_{\text{psr}}^{(\ell)}$ and $f_{\text{sml}}^{(\ell)}$ have range $[-1, 1]$ for each $\ell \in [R]$;
- (iv) $\text{rank } \mathfrak{B} \geq r(\|\mathfrak{B}\|)$;
- (v) $\|f_{\text{sml}}^{(\ell)}\|_2 < \theta$ for each $\ell \in [R]$;
- (vi) $\|\mathfrak{B}\| \leq C_{\text{reg}''}(p, d, R, C_0, \theta, \eta, r)$ and $\deg \mathfrak{B} \leq d$.

Proof. Set $M := \lceil R\theta^{-2} \rceil$. Define nondecreasing functions $r_i: \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ for each $i = 0, \dots, M$ such that $r_0 = r$ and $r_{i+1}(N) \geq r_{\text{reg}'}(p, d, R, N, \eta(N), r_i)$ and such that $r_{i+1}(N) \geq r_i(N)$ for all $i = 0, \dots, M-1$ and all $N \in \mathbb{Z}_{>0}$. Define $r_{\text{reg}''}(p, d, R, C_0, \theta, \eta, r) := r_M(C_0)$.

We construct a list of polynomial factors $\mathfrak{B}_0, \mathfrak{B}_1, \dots, \mathfrak{B}_m$ on V such that

- \mathfrak{B}_i refines \mathfrak{B}_{i-1} for $i = 1, \dots, m$;
- $\text{rank } \mathfrak{B}_i \geq r_{M-i}(\|\mathfrak{B}_i\|)$ for $i = 0, \dots, m$;
- $\|\mathfrak{B}_i\| \leq C_{\text{reg}'}(p, d, R, \|\mathfrak{B}_{i-1}\|, \eta(\|\mathfrak{B}_{i-1}\|), r_{M-i})$ and $\deg \mathfrak{B}_i \leq d$ for $i = 1, \dots, m$.

Suppose we have constructed polynomial factors $\mathfrak{B}_0, \dots, \mathfrak{B}_i$ with the above properties. If $i \geq 1$ and $\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 - \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i-1}]\|_2^2 < \theta^2$ for each $\ell \in [R]$ we halt the iteration.

Otherwise let \mathfrak{B}_{i+1} be the polynomial factor produced by applying Lemma 4.3 to \mathfrak{B}_i with parameters $p, d, R, \|\mathfrak{B}_i\|, \eta(\|\mathfrak{B}_i\|), r_{M-i-1}$. Note that

$$\begin{aligned} \text{rank } \mathfrak{B}_i &\geq r_{M-i}(\|\mathfrak{B}_i\|) \\ &\geq r_{\text{reg}'}(p, d, R, \|\mathfrak{B}_i\|, \eta(\|\mathfrak{B}_i\|), r_{M-i-1}), \end{aligned}$$

so the hypotheses of Lemma 4.3 are satisfied.

Next we claim that this iteration must stop after at most M steps. This is obvious since

$$\sum_{\ell=1}^R \left\| \mathbb{E}[f^{(\ell)} | \mathfrak{B}_i] \right\|_2^2 - \sum_{\ell=1}^R \left\| \mathbb{E}[f^{(\ell)} | \mathfrak{B}_{i-1}] \right\|_2^2 \geq \theta^2$$

for $i = 1, \dots, m$ and the sum is bounded between 0 and R .

Thus we have produced \mathfrak{B}_{m-1} with $m \leq M$ such that \mathfrak{B}_{m-1} refines \mathfrak{B}_0 and $\text{rank } \mathfrak{B}_{m-1} \geq r_{M-m+1}(\|\mathfrak{B}_{m-1}\|) \geq r(\|\mathfrak{B}_{m-1}\|)$ and $\|f^{(\ell)} - \mathbb{E}[f^{(\ell)} | \mathfrak{B}_m]\|_{U^{d+1}} < \eta(\|\mathfrak{B}_{m-1}\|)$ for each $\ell \in [R]$ and $\|\mathbb{E}[f^{(\ell)} | \mathfrak{B}_m] - \mathbb{E}[f^{(\ell)} | \mathfrak{B}_{m-1}]\|_2 < \theta$ for each $\ell \in [R]$. Defining $f_{\text{str}}^{(\ell)} := \mathbb{E}[f^{(\ell)} | \mathfrak{B}_{m-1}]$ and $f_{\text{psr}}^{(\ell)} := f^{(\ell)} - \mathbb{E}[f^{(\ell)} | \mathfrak{B}_m]$ and $f_{\text{sml}}^{(\ell)} := \mathbb{E}[f^{(\ell)} | \mathfrak{B}_m] - \mathbb{E}[f^{(\ell)} | \mathfrak{B}_{m-1}]$, we immediately see that conclusions (i), (ii), (iii), (iv), and (v) hold. Conclusion (vi) holds by defining $C_{\text{reg}'}(p, d, R, C_0, \theta, \eta, r)$ to be the M -fold iteration of the function $N \mapsto C_{\text{reg}'}(p, d, R, N, \eta(N), r_M)$ applied to C_0 . \square

Now we are ready to prove the final arithmetic regularity lemma. In this result we produce a pair of polynomial factors, \mathfrak{B} refined by \mathfrak{B}' . This result is stronger in the sense that the θ that appeared in Lemma 4.4(v) and the d in Lemma 4.4(ii) are allowed to be functions of the size of \mathfrak{B} . Furthermore we have an additional property, Lemma 4.5(vi), that states that a typical subatom a of \mathfrak{B}' “looks like” the atom $\pi(a)$ of \mathfrak{B} that it is contained in.

LEMMA 4.5 (strong arithmetic regularity). *Fix a prime p , positive integers R, C_0, d_0 , a parameter $\zeta > 0$, nonincreasing functions $\eta, \theta: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow (0, 1)$, and nondecreasing functions $d, r: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$. There exist constants $C_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$ and $D_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$ and $r_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space and let \mathfrak{B}_0 be a polynomial factor on V satisfying $\|\mathfrak{B}_0\| \leq C_0$ and $\deg \mathfrak{B}_0 \leq d_0$ and $\text{rank } \mathfrak{B}_0 \geq r_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$. Given functions $f^{(1)}, \dots, f^{(R)}: V \rightarrow [0, 1]$, there exist a polynomial factor \mathfrak{B} and a refinement \mathfrak{B}' both on V with parameters I and I' with the following properties. There exists a decomposition*

$$f^{(\ell)} = f_{\text{str}}^{(\ell)} + f_{\text{psr}}^{(\ell)} + f_{\text{sml}}^{(\ell)}$$

for each $\ell \in [R]$ such that

- (i) $f_{\text{str}}^{(\ell)} = \mathbb{E}[f^{(\ell)} | \mathfrak{B}']$ for each $\ell \in [R]$;
- (ii) $\|f_{\text{psr}}^{(\ell)}\|_{U^{d(\deg \mathfrak{B}, \|\mathfrak{B}\|)+1}} < \eta(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$ for each $\ell \in [R]$;
- (iii) $f_{\text{str}}^{(\ell)}$ and $f_{\text{str}}^{(\ell)} + f_{\text{sml}}^{(\ell)}$ have range $[0, 1]$ and $f_{\text{psr}}^{(\ell)}$ and $f_{\text{sml}}^{(\ell)}$ have range $[-1, 1]$ for each $\ell \in [R]$;
- (iv) $\text{rank } \mathfrak{B} \geq r(\deg \mathfrak{B}, \|\mathfrak{B}\|)$ and $\text{rank } \mathfrak{B}' \geq r(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$;
- (v) $\|f_{\text{sml}}^{(\ell)}\|_2 < \theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)$ for each $\ell \in [R]$;
- (vi) for all but at most a ζ -fraction of $a \in A_{I'}$ it holds that

$$\left| \mathbb{E}_{x \in \mathfrak{B}'^{-1}(a)}[f^{(\ell)}(x)] - \mathbb{E}_{x \in \mathfrak{B}^{-1}(\pi(a))}[f^{(\ell)}(x)] \right| < \zeta$$

for each $\ell \in [R]$ (recall the definition of the atom indexing sets from (3.3) and the projection map $\pi: A_{I'} \rightarrow A_I$ from (3.4));

- (vii) $\|\mathfrak{B}'\| \leq C_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$ and $\deg \mathfrak{B}' \leq D_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r)$.

Proof. Set $M := \lceil R\zeta^{-3} \rceil$. Define nondecreasing functions $r_i: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ for each $i = 0, \dots, M$ such that $r_0 = r$ and $r_{i+1}(D, N) \geq r_{\text{reg}''}(p, d(D, N), R, N, \theta(D, N), \eta(d(D, N), \cdot), r_i(d(D, N), \cdot))$ and such that $r_{i+1}(D, N) \geq r_i(D, N)$ for all $i = 0, \dots, M-1$ and all $D, N \in \mathbb{Z}_{>0}$. Define $r_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r) := r_M(d_0, C_0)$.

We construct a list of polynomial factors $\mathfrak{B}_0, \mathfrak{B}_1, \dots, \mathfrak{B}_m$ on V such that

- \mathfrak{B}_i refines \mathfrak{B}_{i-1} for $i = 1, \dots, m$;
- $\text{rank } \mathfrak{B}_i \geq r_{M-i}(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|)$ for $i = 0, \dots, m$;
- $\|\mathfrak{B}_i\| \leq C_{\text{reg}''}(p, d(\deg \mathfrak{B}_{i-1}, \|\mathfrak{B}_{i-1}\|), R, \|\mathfrak{B}_{i-1}\|, \theta, \eta, r_{M-i}(d(\deg \mathfrak{B}_{i-1}, \|\mathfrak{B}_{i-1}\|), \cdot))$ and $\deg \mathfrak{B}_i \leq d(\deg \mathfrak{B}_{i-1}, \|\mathfrak{B}_{i-1}\|)$ for $i = 1, \dots, m$.

Suppose we have constructed polynomial factors $\mathfrak{B}_0, \dots, \mathfrak{B}_i$ with the above properties. If $i \geq 1$ and $\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_i]\|_2^2 - \|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i-1}]\|_2^2 < \zeta^3$ for each $\ell \in [R]$ we halt the iteration. Otherwise let \mathfrak{B}_{i+1} be the polynomial factor produced by applying Lemma 4.4 to \mathfrak{B}_i with parameters $p, d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), R, \|\mathfrak{B}_i\|, \theta(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \eta(d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \cdot), r_{M-i-1}(d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \cdot)$. Note that

$$\begin{aligned} \text{rank } \mathfrak{B}_i &\geq r_{M-i}(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|) \\ r_{\text{reg}''}(p, d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), R, \|\mathfrak{B}_i\|, \theta(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \eta(d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \cdot), \\ &\quad r_{M-i-1}(d(\deg \mathfrak{B}_i, \|\mathfrak{B}_i\|), \cdot)), \end{aligned}$$

so the hypotheses of Lemma 4.4 are satisfied.

Next we claim that this iteration must stop after at most M steps. This is obvious since

$$\sum_{\ell=1}^R \left\| \mathbb{E}[f^{(\ell)}|\mathfrak{B}_i] \right\|_2^2 - \sum_{\ell=1}^R \left\| \mathbb{E}[f^{(\ell)}|\mathfrak{B}_{i-1}] \right\|_2^2 \geq \zeta^3$$

for $i = 1, \dots, m$ and the sum is bounded between 0 and R .

Thus we have produced \mathfrak{B}_{m-1} and \mathfrak{B}_m with $m \leq M$ such that \mathfrak{B}_{m-1} refines \mathfrak{B}_0 and \mathfrak{B}_m refines \mathfrak{B}_{m-1} . Furthermore, $\text{rank } \mathfrak{B}_{m-1} \geq r_{M-m+1}(\deg \mathfrak{B}_{m-1}, \|\mathfrak{B}_{m-1}\|) \geq r(\deg \mathfrak{B}_{m-1}, \|\mathfrak{B}_{m-1}\|)$ and $\text{rank } \mathfrak{B}_m \geq r_{M-m}(\deg \mathfrak{B}_m, \|\mathfrak{B}_m\|) \geq r(\deg \mathfrak{B}_m, \|\mathfrak{B}_m\|)$. Also $\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_m] - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_{m-1}]\|_2^2 < \zeta^3$ for each $\ell \in [R]$. Let $f^{(\ell)} = f_{\text{str}}^{(\ell)} + f_{\text{psr}}^{(\ell)} + f_{\text{smI}}^{(\ell)}$ be the decomposition produced by the last application of Lemma 4.4. This decomposition satisfies conclusions (i), (ii), (iii), and (v). Conclusion (iv) we already verified, and conclusion (vi) follows from Markov's inequality applied to the bound $\|\mathbb{E}[f^{(\ell)}|\mathfrak{B}_m] - \mathbb{E}[f^{(\ell)}|\mathfrak{B}_{m-1}]\|_2^2 < \zeta^3$. Finally conclusion (vii) holds where we define the pair $(D_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r), C_{\text{reg}'''}(p, R, C_0, d_0, \zeta, \eta, \theta, d, r))$ to be the M -fold iteration of the function $(D, N) \mapsto (C_{\text{reg}''}(p, d(D, N), R, N, \theta, \eta, r_M(d(D, N), \cdot)), d(D, N))$ applied to (d_0, C_0) . \square

4.2. Subatom selection. Let us pause to recall our setup. We have a polynomial factor \mathfrak{B} and a refinement \mathfrak{B}' . The atoms of \mathfrak{B} are indexed by A_I and the atoms of \mathfrak{B}' , which we call “subatoms,” are indexed by \mathfrak{B}' . The map $\pi: A_{I'} \rightarrow A_I$, defined by (3.4), maps a subatom a to the atom $\pi(a)$ which contains it.

The goal of this section is to produce a function $s: A_I \rightarrow A_{I'}$, called a subatom selection function, which for each atom $a \in A_I$ selects a single subatom $s(a) \in A_{I'}$. In

Definition 3.24 we defined subatom selection functions to have a very specific form. In this section we will prove that selecting a uniform random subatom selection function will have many desirable properties with positive probability.

The concept of subatom selection has appeared in past works [8, 7]. However, as those works were restricted to translation-invariant patterns, their subatom selection functions are allowed to be affine in some sense. In fact, in those papers the subatom selection function was essentially chosen to be a uniform random constant function. In our case our subatom selection function must be linear in an appropriate sense. This constraint is why the form of our subatom selection functions is more complicated.

For an atom $a \in A_I$, we write $a_{d,k} \in \mathbb{U}_{k+1}^{I_{d,k}}$ to be the degree d , depth k part of a . Say that a is “good” if $a_{1,0} \neq 0$. Our subatom selection function will be chosen so that $s(a)$ is a regular subatom for every good a : see Theorem 4.6(v) below. For bad a , we cannot ensure that $s(a)$ is regular; this is the set of atoms which we will later apply patching to. For an intuitive explanation of the good/bad distinction, note from Definition 3.24 that if a is bad, then $s(a)$ is constant as s varies. We will prove that if a is good, when s is chosen uniformly random, then $s(a)$ is uniformly distributed over $\pi^{-1}(a)$.

THEOREM 4.6 (subatom selection). *Fix a prime p , positive integers R, c_0 , a parameter $\zeta > 0$, nonincreasing functions $\eta, \theta: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow (0, 1)$, and nondecreasing functions $d, r: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$. There exist constants $C_{\text{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r)$ and $D_{\text{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r)$ and $n_{\text{reg}}(p, c_0, \zeta)$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space satisfying $\dim V \geq n_{\text{reg}}(p, c_0, \zeta)$. Given functions $f^{(1)}, \dots, f^{(R)}: V \rightarrow [0, 1]$, there exist a polynomial factor \mathfrak{B} and a refinement \mathfrak{B}' both on V with parameters I and I' with the following properties. There exists a subatom selection function $s: A_I \rightarrow A_{I'}$ and a decomposition*

$$f^{(\ell)} = f_{\text{str}}^{(\ell)} + f_{\text{psr}}^{(\ell)} + f_{\text{sml}}^{(\ell)}$$

for each $\ell \in [R]$ such that

- (i) $f_{\text{str}}^{(\ell)} = \mathbb{E}[f^{(\ell)} | \mathfrak{B}']$ for each $\ell \in [R]$;
 - (ii) $\|f_{\text{psr}}^{(\ell)}\|_{U^{d(\deg \mathfrak{B}, \|\mathfrak{B}\|)+1}} < \eta(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$ for each $\ell \in [R]$;
 - (iii) $f_{\text{str}}^{(\ell)}$ and $f_{\text{str}}^{(\ell)} + f_{\text{sml}}^{(\ell)}$ have range $[0, 1]$ and $f_{\text{psr}}^{(\ell)}$ and $f_{\text{sml}}^{(\ell)}$ have range $[-1, 1]$ for each $\ell \in [R]$;
 - (iv) $\text{rank } \mathfrak{B} \geq r(\deg \mathfrak{B}, \|\mathfrak{B}\|)$ and $\text{rank } \mathfrak{B}' \geq r(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$;
1. for each $a \in A_I$ with $a_{1,0} \neq 0$, it holds that

$$\|f_{\text{sml}}^{(\ell)} 1_{\mathfrak{B}'^{-1}(s(a))}\|_2 < \theta(\deg \mathfrak{B}, \|\mathfrak{B}\|) \|1_{\mathfrak{B}'^{-1}(s(a))}\|_2$$

for each $\ell \in [R]$;

- (v) for all but at most a ζ -fraction of $a \in A_I$ it holds that

$$\left| \mathbb{E}_{x \in \mathfrak{B}^{-1}(a)}[f^{(\ell)}(x)] - \mathbb{E}_{x \in \mathfrak{B}'^{-1}(s(a))}[f^{(\ell)}(x)] \right| < \zeta$$

for each $\ell \in [R]$;

- (vi) $I_{1,0} \geq c_0$;
- (vii) $\|\mathfrak{B}'\| \leq C_{\text{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r)$ and $\deg \mathfrak{B}' \leq D_{\text{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r)$.

Proof. Define $n_{\text{reg}}(p, c_0, \zeta) := \max\{c_0, \lceil \log_p(2/\zeta) \rceil\}$. Let \mathfrak{B}_0 be a polynomial factor on V defined by $n_{\text{reg}}(p, c_0, \zeta)$ linearly independent linear functions. Define $\theta': \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ by $\theta'(D, N) := \theta(D, N)/(2\sqrt{RN})$. We apply Lemma 4.5 to \mathfrak{B}_0

with parameters $p, R, \|\mathfrak{B}_0\|, 1, \zeta/4, \eta, \theta', d, r$. Let \mathfrak{B} and \mathfrak{B}' be the polynomial factors produced. We immediately have conclusions (i), (ii), (iii), and (iv). Conclusion (vii) follows since \mathfrak{B} refines \mathfrak{B}_0 , which is defined by at least c_0 linear forms. Conclusion (viii) holds by defining $C_{\text{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r) = C_{\text{reg}'''}(p, R, n_{\text{reg}}(p, c_0, \zeta), 1, \zeta/4, \eta, \theta', d, r)$ and $D_{\text{reg}}(p, R, c_0, \zeta, \eta, \theta, d, r) = D_{\text{reg}'''}(p, R, p^{n_{\text{reg}}(p, c_0, \zeta)}, 1, \zeta/4, \eta, \theta', d, r)$.

Let $c_{d,k}^{i,j} \in \mathbb{Z}/p^{k+1}\mathbb{Z}$ be elements chosen independently and uniformly at random for each $(d, k) \in D_p$ and $i \in [I_{1,0}]$ and $I_{d,k} < j \leq I'_{d,k}$. Consider the subatom selection function $s_c: A_I \rightarrow A_{I'}$ defined in Definition 3.24. We claim that with positive probability, this s_c satisfies conclusions (v) and (vi).

Fix $a \in A_I$ with $a_{1,0} \neq 0$. We first claim that for this fixed a , as c varies, the subatom $s(a)$ is uniformly distributed over $\pi^{-1}(a) \subset A_{I'}$. To see this, first note that the univariate homogeneous nonclassical polynomials $P_{d,k}: \mathbb{F}_p \rightarrow \mathbb{U}_{k+1}$ used in the definition of s_c satisfy $P_{d,k}(x) \notin \mathbb{U}_k$ for all $x \neq 0$. This follows by homogeneity: if $P_{d,k}(x) \in \mathbb{U}_k$ for some $x \neq 0$, then $P_{d,k}$ always takes values in \mathbb{U}_k , contradicting the assumption that $P_{d,k}$ has depth exactly k . Thus for $a \in A_I$ with $a_{1,0} \neq 0$, we find that the vector $(P_{d,k}(|a_{1,0}^1|), \dots, P_{d,k}(|a_{1,0}^{I_{1,0}}|)) \in \mathbb{U}_{k+1}^{I_{1,0}}$ does not lie in $\mathbb{U}_k^{I_{1,0}}$.

Considering the definition of s_c , we see that the vector $([s_c]_{d,k}^i)_{I_{d,k} < i \leq I'_{d,k}}$ is produced by the following matrix multiplication:

$$\begin{pmatrix} [s_c]_{d,k}^{I_{d,k}+1} \\ [s_c]_{d,k}^{I_{d,k}+2} \\ \vdots \\ [s_c]_{d,k}^{I'_{d,k}} \end{pmatrix} = \begin{pmatrix} c_{d,k}^{I_{d,k}+1,1} & c_{d,k}^{I_{d,k}+1,2} & \dots & c_{d,k}^{I_{d,k}+1,I_{d,k}} \\ c_{d,k}^{I_{d,k}+2,1} & c_{d,k}^{I_{d,k}+2,2} & \dots & c_{d,k}^{I_{d,k}+2,I_{d,k}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{d,k}^{I'_{d,k},1} & c_{d,k}^{I'_{d,k},2} & \dots & c_{d,k}^{I'_{d,k},I_{d,k}} \end{pmatrix} \begin{pmatrix} P_{d,k}(|a_{1,0}^1|) \\ P_{d,k}(|a_{1,0}^2|) \\ \vdots \\ P_{d,k}(|a_{1,0}^{I_{d,k}}|) \end{pmatrix}.$$

As stated above, the vector on the right lies in $\mathbb{U}_{k+1}^{I_{1,0}}$ but not in $\mathbb{U}_k^{I_{1,0}}$ while the matrix is uniform random with entries in $\mathbb{Z}/p^{k+1}\mathbb{Z}$. This implies that the vector on the left is uniformly distributed over $\mathbb{U}_{k+1}^{I'_{d,k}-I_{d,k}}$, as desired. This holds for each $(d, k) \in D_p$ and furthermore, since for each (d, k) the c matrices are chosen independently, one can see that the resulting vectors are also independent, proving the desired result.

Now note that by Lemma 4.5(vi), for each $\ell \in [R]$,

$$\sum_{a \in A_{I'}} \|f_{\text{sm}}^{(\ell)} 1_{\mathfrak{B}'^{-1}(a)}\|_2^2 \leq \frac{\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)^2}{4R\|\mathfrak{B}\|^2} \sum_{a \in A_{I'}} \|1_{\mathfrak{B}'^{-1}(a)}\|_2^2.$$

Thus by Markov's inequality, for $a' \in A_{I'}$ chosen uniformly at random, with probability at least $1 - 1/(4R\|\mathfrak{B}\|^2)$ the following holds:

$$(4.1) \quad \|f_{\text{sm}}^{(\ell)} 1_{\mathfrak{B}'^{-1}(a')}\|_2^2 \leq \theta(\|\mathfrak{B}\|)^2 \|1_{\mathfrak{B}'^{-1}(a')}\|_2^2.$$

Thus for a fixed $a \in A_I$ with $a_{1,0} \neq 0$, we have that for c chosen at random, we have $s_c(a)$ satisfies (4.1) with probability at least $1 - 1/(4R\|\mathfrak{B}\|)$. Taking a union bound over all choices of $a \in A_I$ and $\ell \in [R]$, we see that with probability at least $3/4$, conclusion (v) holds.

Now we deduce conclusion (vi). First note that the fraction of $a \in A_I$ which satisfy $a_{1,0} = 0$ is $p^{-I_{1,0}} \leq \zeta/2$. For the other $a \in A_I$, the expected fraction of $a \in A_I$ that fail to satisfy the desired inequality is at most $\zeta/4$. Thus by Markov's inequality, with probability at least $1/2$, at most $\zeta/2$ fraction of $a \in A_I$ satisfy $a_{1,0} \neq 0$ and fail to satisfy the desired inequality. Thus with probability at least $1/2$, conclusion (vi) holds. \square

5. Patching. In this section we prove the main patching theorem we need, Theorem 5.9. We call this result a supersaturation dichotomy theorem for reasons we will explain shortly. This is proved from a Ramsey dichotomy result, Theorem 5.6. This Ramsey dichotomy result is fairly easy to deduce from an appropriate Ramsey-type theorem which we also prove in this section.

5.1. A motivating example. To motivate the kind of results proved in this section, we give a series of dichotomy results in the much simpler setting of edge-colored graphs. These results are analogous to the dichotomy results we will prove later in this section.

We start with a tautological dichotomy which is essentially trivial.

FACT 0 (tautological dichotomy). *Let \mathcal{H} be a finite set of red/blue edge-colored graphs. There exists an integer $n_0 = n_0(\mathcal{H})$ such that the following holds. Either*

- (a) *for every n , there exists a red/blue edge-coloring of K_n that contains no subgraph from \mathcal{H} , or*
- (b) *every red/blue edge-coloring of K_n with $n \geq n_0$ contains a subgraph from \mathcal{H} .*

In case (a), we find \mathcal{H} -free colorings of arbitrarily large cliques. Using Ramsey's theorem, we can find arbitrarily large monochromatic cliques in these \mathcal{H} -free colorings. This lets us boost the previous result into the following stronger Ramsey dichotomy result. (One could also prove this result directly from Ramsey's theorem.)

FACT 1 (Ramsey dichotomy). *Let \mathcal{H} be a finite set of red/blue edge-colored graphs. There exists an integer $n_0 = n_0(\mathcal{H})$ such that the following holds. Either*

- (a) *the all-red coloring of K_n or the all-blue coloring of K_n contains no subgraph from \mathcal{H} for every n , or*
- (b) *every red/blue edge-coloring of K_n with $n \geq n_0$ contains a subgraph from \mathcal{H} .*

A sampling argument allows one to boost this dichotomy result into a supersaturation dichotomy result. This result allows one to find many copies of the desired graph instead of just one.

FACT 2 (supersaturation dichotomy). *Let \mathcal{H} be a finite set of red/blue edge-colored graphs. There exist $n_0 = n_0(\mathcal{H})$ and $\epsilon = \epsilon(\mathcal{H}) > 0$ such that the following holds. Either*

- (a) *the all-red coloring of K_n or the all-blue coloring of K_n contains no subgraph from \mathcal{H} for every n , or*
- (b) *for every red/blue edge-coloring of K_n with $n \geq n_0$ there exists a subgraph $H \in \mathcal{H}$ which has density at least ϵ in the coloring.*

The main dichotomy result which we will prove is analogous to Fact 1 and the main patching result is analogous to Fact 2.

5.2. Colored labeled patterns and canonical colorings. In this section we define the setting in which we prove our dichotomy results. Unfortunately for our application it is not sufficient to study colored patterns in colorings $f: V \rightarrow \mathcal{S}$. In our setting we also have a polynomial factor \mathfrak{B} on V and it is important to know how the pattern interacts with the factor. To do so we introduce the notion of a colored labeled pattern.

DEFINITION 5.1. *For a prime p , a finite set \mathcal{S} , and a parameter list $I \in \mathcal{I}_p$, an \mathcal{S} -colored I -labeled pattern over \mathbb{F}_p consisting of m linear forms in ℓ variables is a triple (\mathbf{L}, ψ, ϕ) given by*

- a system $\mathbf{L} = (L_1, \dots, L_m)$ of m linear forms in ℓ variables,
- a coloring $\psi: [m] \rightarrow \mathcal{S}$, and
- a labeling $\phi: [m] \rightarrow A_I$ (recall the definition of the atom-indexing set A_I from (3.3)).

Given a finite-dimensional \mathbb{F}_p -vector space V , a function $f: V \rightarrow \mathcal{S}$, and a polynomial factor \mathfrak{B} on V with parameters I , an (\mathbf{L}, ψ, ϕ) -instance in (f, \mathfrak{B}) is some $\mathbf{x} \in V^\ell$ such that $f(L_i(\mathbf{x})) = \psi(i)$ for all $i \in [m]$ and $\mathfrak{B}(L_i(\mathbf{x})) = \phi(i)$ for all $i \in [m]$. An instance is called generic if x_1, \dots, x_ℓ are linearly independent.

DEFINITION 5.2. For an \mathcal{S} -colored I -labeled pattern (\mathbf{L}, ψ, ϕ) consisting of m linear forms, a finite-dimensional \mathbb{F}_p -vector space V , a function $f: V \rightarrow \mathcal{S}$, and a polynomial factor \mathfrak{B} on V with parameters I , define the (\mathbf{L}, ψ, ϕ) -density in (f, \mathfrak{B}) to be

$$\Lambda_{\mathbf{L}}(1_{f^{-1}(\psi(1)) \cap \mathfrak{B}^{-1}(\phi(1))}, \dots, 1_{f^{-1}(\psi(m)) \cap \mathfrak{B}^{-1}(\phi(m))}).$$

Given a set $X \subseteq V$, define the relative density of (\mathbf{L}, ψ, ϕ) in X to be

$$\frac{\Lambda_{\mathbf{L}}(f_1, \dots, f_m)}{\Lambda_{\mathbf{L}}(1_X, \dots, 1_X)},$$

where $f_i := 1_{X \cap f^{-1}(\psi(i)) \cap \mathfrak{B}^{-1}(\phi(i))}$.

In the graph example we gave in subsection 5.1, the all-red and all-blue colorings played a special role. In this setting the analogous role is played by so-called canonical colorings which we define below. A similar concept appears in the simpler application of patching given by the authors and Fox [12]. In fact, similar ideas can be extracted from classical results in Ramsey theorem such as Rado's theorem.

DEFINITION 5.3. Define the first nonzero coordinate function $\text{fnz}: \mathbb{F}_p^n \rightarrow \mathbb{F}_p$ by $\text{fnz}(0, \dots, 0) := 0$ and $\text{fnz}(x_1, \dots, x_n) := x_k$ where $x_1 = \dots = x_{k-1} = 0$ and $x_k \neq 0$. Given a finite-dimensional \mathbb{F}_p -vector space V equipped with an isomorphism $\iota: V \xrightarrow{\sim} \mathbb{F}_p^n$, define the function $\text{fnz}_\iota: V \rightarrow \mathbb{F}_p$ by $\text{fnz}_\iota(x) := \text{fnz}(\iota(x))$.

DEFINITION 5.4. Fix a prime p , a finite set \mathcal{S} , a parameter list $I \in \mathcal{I}_p$, and a function $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$. For a finite-dimensional \mathbb{F}_p -vector space V equipped with an isomorphism $\iota: V \xrightarrow{\sim} \mathbb{F}_p^n$ and a polynomial factor \mathfrak{B} on V with parameters I , define the ξ -canonical coloring $\Xi_{\xi, \iota, \mathfrak{B}}: V \rightarrow \mathcal{S}$ by $\Xi_{\xi, \iota, \mathfrak{B}}(x) := \xi(\text{fnz}_\iota(x), \mathfrak{B}(x))$. Furthermore, if \mathcal{S} is equipped with an \mathbb{F}_p^\times -action, say that ξ is projective if the same is true for every function $\Xi_{\xi, \iota, \mathfrak{B}}$. (Note that this property is equivalent to the condition that ξ preserves the action of \mathbb{F}_p^\times , i.e., $\xi(cx, c \cdot a) = c \cdot \xi(x, a)$ for all $c \in \mathbb{F}_p^\times$, all $x \in \mathbb{F}_p$, and all $a \in A_I$. Recall the action of \mathbb{F}_p^\times on A_I defined in (3.5).)

We will soon prove that if there exists a generic H -instance in some ξ -canonical coloring $\Xi_{\xi, \iota, \mathfrak{B}}$ of some \mathbb{F}_p -vector space, then there exists an H -instance in every ξ -canonical coloring of every sufficiently large vector space equipped with a polynomial factor of sufficiently high rank. We say that ξ canonically induces H if the former condition holds.

DEFINITION 5.5. Given a prime p , a finite set \mathcal{S} , a parameter list $I \in \mathcal{I}_p$, a function $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$, and an \mathcal{S} -colored I -labeled pattern $H = (\mathbf{L}, \psi, \phi)$, say that ξ canonically induces H if the following holds. There exists some $n \geq 0$ and a polynomial factor \mathfrak{B} on \mathbb{F}_p^n with parameters I such that there exists a generic H -instance in $(\Xi_{\xi, \text{Id}, \mathfrak{B}}, \mathfrak{B})$. For a finite set of \mathcal{S} -colored I -labeled patterns \mathcal{H} , say that ξ canonically induces \mathcal{H} if ξ canonically induces some $H \in \mathcal{H}$.

5.3. Ramsey dichotomy result. In this section we prove our Ramsey dichotomy result, Theorem 5.6. To do so we will need two Ramsey-type theorems for this setting, Lemmas 5.7 and 5.8.

Ramsey-type results, specifically the affine Ramsey theorem of Graham and Rothschild, allow one to find a large monochromatic *affine* subspace in a coloring $f: V \rightarrow \mathcal{S}$. In our setting we have a polynomial factor \mathfrak{B} on V . The main technical result of this section, Lemma 5.8, is analogous to the affine Ramsey theorem with the additional restriction that the affine subspace U produced respects the structure of \mathfrak{B} (essentially that $\mathfrak{B}|_U$ has sufficiently high rank).

From this result, a fairly elementary argument lets us deduce Lemma 5.7, which says that one can find a large *linear* subspace which is canonically colored. Using this Ramsey result we can easily deduce the main Ramsey dichotomy result, Theorem 5.6. One has to be careful with some technical details related to high rank. For example, we will need to check that there exists at least one \mathbf{L} -instance lying in the correct atoms of \mathfrak{B} .

THEOREM 5.6 (Ramsey dichotomy). *Fix a prime p , a finite set \mathcal{S} with an \mathbb{F}_p^\times -action, a parameter list $I \in \mathcal{I}_p$, and a positive integer ℓ_0 . There exist constants $n_{\text{dich}} = n_{\text{dich}}(p, |\mathcal{S}|, I, \ell_0)$ and $r_{\text{dich}} = r_{\text{dich}}(p, |\mathcal{S}|, I, \ell_0)$ such that the following holds. Let \mathcal{H} be a finite set of \mathcal{S} -colored, I -labeled patterns each defined by a system of linear forms in at most ℓ_0 variables. Either*

- (a) *there exists a projective $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$ that does not canonically induce \mathcal{H} , or*
- (b) *for every finite-dimensional \mathbb{F}_p -vector space V satisfying $\dim V \geq n_{\text{dich}}$, every projective function $f: V \rightarrow \mathcal{S}$, and every polynomial factor \mathfrak{B} on V with parameters I which has rank at least r_{dich} , there is a generic H -instance in (f, \mathfrak{B}) for some $H \in \mathcal{H}$.*

Proof. Define m to be the smallest positive integer such that the following holds. Let H be an \mathcal{S} -colored, I -labeled pattern defined by a system of linear forms in at most ℓ_0 variables and let $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$ be a projective function. If ξ canonically induces H , then there exists some $n_H \leq m$ and a polynomial factor \mathfrak{B}_H on $\mathbb{F}_p^{n_H}$ with parameters I such that there exists a generic H -instance in $(\Xi_{\xi, \text{Id}, \mathfrak{B}_H}, \mathfrak{B}_H)$. This is well defined since there are only a finite number of \mathcal{S} -colored, I -labeled patterns defined by a system of linear forms in at most ℓ_0 variables.

LEMMA 5.7. *Fix a prime p , a finite set \mathcal{S} , a parameter list $I \in \mathcal{I}_p$, and positive integers m, r_0, n_0 . There exist constants $n_{\text{ramsey}} = n_{\text{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0)$ and $r_{\text{ramsey}} = r_{\text{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0)$ such that the following holds. Let V be a finite dimensional \mathbb{F}_p -vector space satisfying $\dim V \geq n_{\text{ramsey}}$, let \mathfrak{B} be a polynomial factor on V with parameters I such that $\text{rank } \mathfrak{B} \geq r_{\text{ramsey}}$, and let $f: V \rightarrow \mathcal{S}$ be a function. Then there exists a subspace $U \leq V$, and linear functions $P_1, \dots, P_m: V \rightarrow \mathbb{F}_p$, and a function $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$ such that the following holds:*

- (i) $\xi(\text{fnz}(P_1(x), \dots, P_m(x)), \mathfrak{B}(x)) = f(x)$ for all $x \in U$ that also satisfy $(P_1(x), \dots, P_m(x)) \neq (0, \dots, 0)$;
- (ii) $\xi(0, 0) = f(0)$;
- (iii) *the polynomial factor \mathfrak{B}' on U defined by the homogeneous nonclassical polynomials that define $\mathfrak{B}|_U$ in addition to the polynomials P_1, \dots, P_m satisfies $\text{rank } \mathfrak{B}' \geq r_0$;*
- (iv) $\dim U \geq n_0$.

Let us show how this lemma completes the proof. Define

$$r_0 := r_{\text{equi}}\left(p, \deg I, \frac{1}{2(p^m \|I\|)^{p^m}}\right) \quad \text{and} \quad n_0 := 2p^m(m + \lceil \log_p \|I\| \rceil).$$

Then define

$$n_{\text{dich}}(p, |\mathcal{S}|, I, \ell_0) := n_{\text{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0)$$

and

$$r_{\text{dich}}(p, |\mathcal{S}|, I, \ell_0) := r_{\text{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0).$$

Let \mathcal{H} be a finite set of \mathcal{S} -colored, I -labeled patterns each defined by a system of linear forms in at most ℓ_0 variables. Suppose (a) does not hold. Thus for every projective $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$, there exists an $H \in \mathcal{H}$ such that ξ canonically induces H .

Now we apply Lemma 5.7 to $f: V \rightarrow \mathcal{S}$. This produces a subspace $U \leq V$, linear functions P_1, \dots, P_m , and a function $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$ with several desirable properties.

First note that since $f: V \rightarrow \mathcal{S}$ is projective, the same is true of $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$. Thus by assumption there exists an $H \in \mathcal{H}$ such that ξ canonically induces H . By the choices in the first paragraph, there exists an $n_H \leq m$ and a polynomial factor \mathfrak{B}_H on $\mathbb{F}_p^{n_H}$ such that $(\Xi_{\xi, \text{Id}, \mathfrak{B}_H}, \mathfrak{B}_H)$ contains a generic H -instance.

To complete the proof, all we need to show is that there exists an injective linear map $\kappa: \mathbb{F}_p^{n_H} \rightarrow U$ such that $\mathfrak{B}_H(x) = \mathfrak{B}(\kappa(x))$ for all $x \in \mathbb{F}_p^{n_H}$ and $\text{fnz}(x) = \text{fnz}(P_1(\kappa(x)), \dots, P_m(\kappa(x)))$ for all $x \in \mathbb{F}_p^{n_H}$. This follows by an application of equidistribution (Theorem 3.19).

Recall the definition of \mathbf{L}^{n_H} (Definition 2.7), the system of p^{n_H} linear forms in n_H variables that define an n_H -dimensional subspace.

Say that \mathfrak{B}' has parameters I' . Thus the atom-indexing set of \mathfrak{B}' can be written as $A_{I'} \simeq \mathbb{F}_p^m \times A_I$. We define the following tuple of atoms $\mathbf{a} = (a_i)_{i \in \mathbb{F}_p^{n_H}}$ by $a_i = ((i_1, \dots, i_{n_H}, 0, \dots, 0), \mathfrak{B}_H(\mathbf{i}))$ where there are $m - n_H$ 0's. We claim that $\mathbf{a} \in \Phi_{I'}(\mathbf{L}^{n_H})$. We can check this separately for the first and second coordinates; each is trivial.

Thus by Theorem 3.19 and the rank bound on \mathfrak{B}' , we have

$$\begin{aligned} & \Pr_{x_1, \dots, x_{n_H} \in U} (\mathfrak{B}'(L_{\mathbf{i}}^{n_H}(x_1, \dots, x_{n_H})) = a_i \text{ for all } \mathbf{i} \in \mathbb{F}_p^{n_H}) \\ & \geq \frac{1}{|\Phi_{I'}(\mathbf{L}^{n_H})|} - \frac{1}{2(p^m \|I\|)^{p^m}} \geq \frac{1}{2(p^m \|I\|)^{p^m}}. \end{aligned}$$

We wish to find a single tuple $(x_1, \dots, x_{n_H}) \in V^{n_H}$ that satisfies the above condition and also has x_1, \dots, x_{n_H} linearly independent. The number of linearly dependent tuples is small, so we calculate that the number of good tuples is at least

$$\frac{|U|^{n_H}}{2(p^m \|I\|)^{p^m}} - |U|^{n_H-1} p^{n_H}.$$

This is positive by our assumption that $\dim U \geq n_0$. Thus there exists some good tuple $(x_1, \dots, x_{n_H}) \in V^{n_H}$. Defining $\kappa: \mathbb{F}_p^{n_H} \rightarrow U$ by $\kappa(\mathbf{i}) := L_{\mathbf{i}}^{n_H}(x_1, \dots, x_{n_H})$ has all the desired properties. Thus we have shown (b) assuming that (a) does not hold. \square

Proof of Lemma 5.7. Define $M := m|\mathcal{S}|^{p\|I\|}$. Our strategy is to find a large subspace U_M and linear functions $P_1, \dots, P_M: V \rightarrow \mathbb{F}_p$ such that for $x \in U_M$, the value of $f(x)$ only depends on $\mathfrak{B}(x)$, $\text{fnz}(P_1(x), \dots, P_M(x))$, and the index k such that $P_1(x) = \dots = P_{k-1}(x) = 0$ and $P_k(x) \neq 0$. Once we have found such a configuration, we can complete the proof by a simple pigeonhole argument.

Define

$$\mathcal{L} := \{p^s P_{(p-1)s+1, s}^i : s \geq 0, i \in [I_{(p-1)s+1, s}]\},$$

where $P_{d,k}^i$ is the i th homogeneous nonclassical polynomial of degree d and depth k defining \mathfrak{B} . Note that \mathcal{L} is a finite set of linear functions (in particular, it is the set of all $p^s P$ that are classical linear polynomials where $s \geq 0$ is a nonnegative integer and P is one of the homogeneous nonclassical polynomials that define \mathfrak{B} .) It is immediate from the definition of rank that if P_1, \dots, P_m are linear functions such that $\{P_1, \dots, P_m\} \cup \mathcal{L}$ are linearly independent, then $\text{rank } \mathfrak{B}' = \text{rank } \mathfrak{B}$. We will use this fact to guarantee conclusion (iii).

Our main tool is the following Ramsey-type lemma which is somewhat analogous to the affine Ramsey theorem of Graham and Rothschild.

LEMMA 5.8. *Fix a prime p , a finite set \mathcal{S} , a parameter list $I \in \mathcal{I}_p$, and positive integers n_0, r_0 . There exist constants $n_{\text{ramsey}'} = n_{\text{ramsey}'}(p, |\mathcal{S}|, I, n_0, r_0)$ and $r_{\text{ramsey}'} = r_{\text{ramsey}'}(p, |\mathcal{S}|, I, n_0, r_0)$ such that the following holds. Let V be a finite-dimensional \mathbb{F}_p -vector space satisfying $\dim V \geq n_{\text{ramsey}'}$ and let $P: V \rightarrow \mathbb{F}_p$ be a nontrivial linear function. Let \mathfrak{B} be a polynomial factor on V with parameters I and let \mathfrak{B}' be the common refinement of \mathfrak{B} and $\{P\}$. Suppose that $\text{rank } \mathfrak{B}' \geq r_{\text{ramsey}'}$. Let $f: V \rightarrow \mathcal{S}$ be a function. Then there exists a subspace $U \leq V$ contained in the zero set of P , a vector $z \in V$ such that $P(z) = 1$, and a function $\chi: A_I \rightarrow \mathcal{S}$ such that the following holds:*

- (i) $\chi(\mathfrak{B}(x)) = f(x)$ for all $x \in z + U$;
- (ii) $\text{rank } \mathfrak{B}|_U \geq r_0$;
- (iii) $\dim U \geq n_0$.

Define r_1, \dots, r_M and n_1, \dots, n_M by

$$n_i := \max\{n_{\text{ramsey}'}(p, |\mathcal{S}|^{p-1}, I, n_{i-1}, r_{i-1}), |\mathcal{L}| + 1\}$$

and

$$r_i := r_{\text{ramsey}'}(p, |\mathcal{S}|^{p-1}, I, n_{i-1}, r_{i-1}).$$

Then define

$$n_{\text{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0) := n_M \quad \text{and} \quad r_{\text{ramsey}}(p, |\mathcal{S}|, I, m, r_0, n_0) := r_M.$$

We will find nested subspaces $V = U_0 \geq U_1 \geq \dots \geq U_M$, linear functions $P_i: V \rightarrow \mathbb{F}_p$, and functions $\xi_i: (\mathbb{F}_p \setminus \{0\}) \times A_I \rightarrow \mathcal{S}$ such that the following holds for each $i \in [M]$:

- $\xi_i(P_i(x), \mathfrak{B}(x)) = f(x)$ for all $x \in U_i$ that satisfy $P_1(x) = \dots = P_{i-1}(x) = 0$ and $P_i(x) \neq 0$;
- $\{P_1, \dots, P_i\} \cup \mathcal{L}$ are linearly independent;
- $\text{rank } \mathfrak{B}|_{W_i} \geq r_{M-i}$ where $W_i := \{x \in U_i : P_1(x) = \dots = P_i(x)\}$;
- $\dim W_i \geq n_{M-i}$ where $W_i := \{x \in U_i : P_1(x) = \dots = P_i(x)\}$.

Suppose we have defined $V = U_0 \geq U_1 \geq \dots \geq U_i$, linear functions $P_1, \dots, P_i: V \rightarrow \mathbb{F}_p$, and functions $\xi_1, \dots, \xi_i: (\mathbb{F}_p \setminus \{0\}) \times A_I \rightarrow \mathcal{S}$ with the above properties.

Define $W := \{x \in U_i : P_1(x) = \dots = P_i(x) = 0\}$. We have $\dim W \geq n_i > |\mathcal{L}|$. Pick an arbitrary $y \in W$ such that $y \neq 0$ but all of the linear functions in \mathcal{L} vanish on y . Let $P_{i+1}: V \rightarrow \mathbb{F}_p$ be an arbitrary linear function such that $P_{i+1}(y) = 1$. Note that automatically we have $\{P_1, \dots, P_i, P_{i+1}\} \cup \mathcal{L}$ are linearly independent.

Define $W' := \{x \in W : P_{i+1}(x) = 0\}$. Note that the subspace W is partitioned into hyperplanes as $W = W' \sqcup (y + W') \sqcup (2y + W') \sqcup \dots$. Write $\overline{\mathcal{S}} := \mathcal{S}^{\mathbb{F}_p^\times}$. Then define $\overline{f}: W \rightarrow \overline{\mathcal{S}}$ by

$$\overline{f}(x + ty) := (f(bx + by))_{b \in \mathbb{F}_p^\times} \quad \text{for } x \in W' \text{ and } t \in \mathbb{F}_p.$$

We apply Lemma 5.8 to $\overline{f}, \mathfrak{B}|_{W, P_{i+1}}$ with parameters n_{M-i-1}, r_{M-i-1} to produce a subspace $U'_i \leq W'$, a vector $z \in y + U'_i$, and a function $\chi: A_I \rightarrow \overline{\mathcal{S}}$ with several desirable properties.

We have $\chi(\mathfrak{B}(x + z)) = \overline{f}(x + z)$ for all $x \in U'_i$. Looking at the b th coordinate of this equation for some $b \in \mathbb{F}_p^\times$ gives $\chi(\mathfrak{B}(x + z))_b = f(bx + bz)$. Finally using the homogeneity of \mathfrak{B} (recall the action of \mathbb{F}_p^\times on A_I defined in (3.5)) gives $\chi(b^{-1} \cdot \mathfrak{B}(bx + bz))_b = f(bx + bz)$ for all $x \in U'_i$ and all $b \in \mathbb{F}_p^\times$.

Define $U_{i+1} \leq U_i$ to be a $(\dim W + i + 1)$ -dimensional subspace of U_i that contains z and W and such that none of P_1, \dots, P_i are identically 0 on U_{i+1} . Then define $\xi_{i+1}: (\mathbb{F}_p \setminus \{0\}) \times A_I \rightarrow \mathcal{S}$ by

$$\xi_{i+1}(b, a) := \chi(b^{-1} \cdot a)_b.$$

Note $\dim U_{i+1} \geq \dim W' \geq m_{i+1}$ and $\text{rank } \mathfrak{B}|_{U_{i+1}} \geq \text{rank } \mathfrak{B}|_{W'} \geq r_{i+1}$. Furthermore, for $x \in U_{i+1}$ such that $P_1(x) = \dots = P_{i-1}(x) = 0$ and $P_i(x) = b \neq 0$, we can write $x = bx' + bz$ for some $x' \in W'$. Then $f(x) = \overline{f}(x' + z)_b = \chi(\mathfrak{B}(x' + z))_b = \chi(b^{-1} \cdot \mathfrak{B}(bx' + bz))_b = \xi_{i+1}(b, \mathfrak{B}(x))$, as desired.

Thus we have defined a sequence of nested subspaces $V = U_0 \leq \dots \leq U_M$, linear functions $P_1, \dots, P_m: V \rightarrow \mathbb{F}_p$, and functions $\xi_1, \dots, \xi_M: (\mathbb{F}_p \setminus \{0\}) \times A_I \rightarrow \mathcal{S}$ with the above properties.

Finally note that the number of possible functions $(\mathbb{F}_p \setminus \{0\}) \times A_I \rightarrow \mathcal{S}$ is at most $|\mathcal{S}|^{p\|I\|}$. Thus by the pigeonhole principle, there exists $1 \leq i_1 < \dots < i_m \leq M$ such that $\xi_1 = \dots = \xi_m$. Define $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$ by $\xi(0, a) = f(0)$ for all $a \in A$ and $\xi(b, a) = \xi_1(b, a)$ for all $b \in (\mathbb{F}_p \setminus \{0\})$ and all $a \in A$. Define $W_M := \{x \in U_M : P_1(x) = \dots = P_M(x) = 0\}$ and let U be a $(\dim W_M + m)$ -dimensional subspace of V that contains W_M and such that none of P_{i_1}, \dots, P_{i_m} are identically 0 on U . Then $U, P_{i_1}, \dots, P_{i_m}$, and ξ have all the desired properties. \square

We now prove our Ramsey-type lemma. We are given a coloring $f: V \rightarrow \mathcal{S}$ and a polynomial factor \mathfrak{B}' . Our goal is to find an affine subspace $z + U$ such that the coloring $f|_{z+U}$ only depends on \mathfrak{B} and additionally $\mathfrak{B}|_U$ has rank at least r_0 .

We satisfy the second condition in a quite crude way. We start by fixing some polynomial factor \mathfrak{B}_1 on $\mathbb{F}_p^{n_1}$ with rank at least r_0 . Then we require $\mathfrak{B}|_U$ to be isomorphic to \mathfrak{B}_1 , which ensures that it has rank at least r_0 .

The proof works by applying the arithmetic regularity lemma, Lemma 4.4, to produce a polynomial factor \mathfrak{B}'' which regularizes the coloring. The arithmetic regularity lemma is a powerful tool that when combined with the equidistribution (Theorem 3.19), allows us to not just find one, but actually count the number of affine subspaces U where each point of U lies in a specified atom of \mathfrak{B}'' and is colored a specified color by f . While there is a fair amount of technical work that goes into applying arithmetic regularity, the high-level overview is that these counts of U are “close to expected,” which allows us to find at least one U with all the desired properties.

Proof of Lemma 5.8. Define constants

$$n_1 := \max\{n_0, n_{\text{high-rank}}(p, I, r_0 + p)\}$$

and

$$\theta := \frac{1}{8(2|\mathcal{S}|)^{2p^{n_1}}} \quad \text{and} \quad \theta' := \frac{\theta}{\|I\|\sqrt{2p|\mathcal{S}|}}.$$

Define the nonincreasing functions $\alpha: \mathbb{Z}_{>0} \rightarrow (0, 1)$ by

$$\alpha(N) := \frac{1}{2N^{2p^{n_1}}}$$

and $\eta: \mathbb{Z}_{>0} \rightarrow (0, 1)$ by

$$\eta(N) := \frac{1}{8(3|\mathcal{S}|N)^{p^{n_1}}},$$

and define the nondecreasing function $r: \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ by

$$r(N) := r_{\text{equi}}(p, p^{n_1}, \alpha(N)).$$

Define

$$n_{\text{ramsey}'}(p, |\mathcal{S}|, I, n_0, r_0) := 2p^{n_1} \lceil \log_p(16|\mathcal{S}|C_{\text{reg}''}(p, p^{n_1}, |\mathcal{S}|, p\|I\|, \theta', \eta, r)) \rceil$$

and

$$r_{\text{ramsey}'}(p, |\mathcal{S}|, I, n_0, r_0) := r_{\text{reg}''}(p, p^{n_1}, |\mathcal{S}|, p\|I\|, \theta', \eta, r).$$

For $c \in \mathcal{S}$, define $f^{(c)}: V \rightarrow [0, 1]$ by $f^{(c)} := 1_{f^{-1}(c)}$. We are now ready to proceed with the proof. We apply the arithmetic regularity lemma, Lemma 4.4, to the polynomial factor \mathfrak{B}' on V and the functions $(f^{(c)})_{c \in \mathcal{S}}$ with parameters $p, p^{n_1}, |\mathcal{S}|, p\|I\|, \theta', \eta, r$. This produces a polynomial factor \mathfrak{B}'' refining \mathfrak{B}' and decompositions $f^{(c)} = f_{\text{str}}^{(c)} + f_{\text{psr}}^{(c)} + f_{\text{sml}}^{(c)}$ with several desirable properties.

Say that \mathfrak{B}' has parameters I' and \mathfrak{B}'' has parameters I'' (note that $\|I'\| = p\|I\|$). We consider the atom-indexing set of \mathfrak{B}'' (see (3.3) for the definition) as $A_{I''} \simeq \mathbb{F}_p \times A_I \times A_{I''-I'}$.

Say that an atom $a \in A_{I''}$ is *regular* if

$$\|f_{\text{sml}}^{(c)} 1_{\mathfrak{B}''^{-1}(a)}\|_2 \leq \theta \|1_{\mathfrak{B}''^{-1}(a)}\|_2 \quad \text{for all } c \in \mathcal{S}.$$

Our first goal is to find $s \in A_{I''-I'}$ such that all atoms of the form $(1, a, s) \in A_{I''} \simeq \mathbb{F}_p \times A_I \times A_{I''-I'}$ are regular.

By Lemma 4.4(v), for each $c \in \mathcal{S}$,

$$\sum_{a \in A_{I''}} \|f_{\text{sml}}^{(c)} 1_{\mathfrak{B}''^{-1}(a)}\|_2^2 \leq \frac{\theta^2}{2p|\mathcal{S}|\|I\|^2} \sum_{a \in A_{I''}} \|1_{\mathfrak{B}''^{-1}(a)}\|_2^2.$$

Thus at least a $(1 - 1/(2p\|I\|^2))$ -fraction of atoms are regular. For each $a \in A_I$, at least a $(1 - 1/(2\|I\|))$ -fraction of the atoms of the form $(1, a, s)$ are regular, for $s \in A_{I''-I'}$. Thus by a union bound there exists some $s \in A_{I''-I'}$ such that $(1, a, s)$ is regular for all $a \in A_I$. Fix this value of s for the rest of the proof.

Define $\chi: A_I \rightarrow \mathcal{S}$ such that $\chi(a)$ is a color that appears in the atom $\mathfrak{B}^{-1}(1, a, s)$ with density at least $1/|\mathcal{S}|$.

By the definition of n_1 and Lemma 3.15, there exists a polynomial factor \mathfrak{B}_1 on $\mathbb{F}_p^{n_1}$ with parameters I and satisfying $\text{rank } \mathfrak{B}_1 \geq r_0 + p$. Our goal is to find vectors $x_0, x_1, \dots, x_{n_1} \in V$ such that

- $\mathfrak{B}''(x_0 + i_1 x_1 + \cdots + i_{n_1} x_{n_1}) = (1, \mathfrak{B}_1(i_1, \dots, i_{n_1}), s)$ for all $(i_1, \dots, i_{n_1}) \in \mathbb{F}_p^{n_1}$;
- $f(x_0 + i_1 x_1 + \cdots + i_{n_1} x_{n_1}) = \chi(\mathfrak{B}_1(i_1, \dots, i_{n_1}))$ for all $(i_1, \dots, i_{n_1}) \in \mathbb{F}_p^{n_1}$;
- x_1, \dots, x_{n_1} are linearly independent.

We choose $x_0, x_1, \dots, x_{n_1} \in V$ independently and uniformly at random. Let p_1 be the probability that this choice of \mathbf{x} satisfies all three conditions above. First note that the probability that x_0, \dots, x_{n_1} are linearly dependent is at most $p^{n_1+1}/|V|$. Let p_2 be the probability that this choice of \mathbf{x} satisfies the first two conditions above. We have shown that

$$p_1 \geq p_2 - p^{n_1+1}/|V|.$$

Let $\mathbf{L} = (L_i)_{i \in \mathbb{F}_p^{n_1}}$ be the system of p^{n_1} linear forms in $n_1 + 1$ variables defined by

$$L_i(x_0, x_1, \dots, x_{n_1}) := x_0 + i_1 x_1 + \cdots + i_{n_1} x_{n_1}.$$

This system defines an n_1 -dimensional affine subspace. One can easily see that \mathbf{L} is finite complexity and in fact its complexity is at most p^{n_1} (see Remark 3.4).

Define $g^{(i)} := f(\chi(\mathfrak{B}_1(i)))$ and define $g_{\text{str}}^{(i)}, g_{\text{sml}}^{(i)}, g_{\text{psr}}^{(i)}$ similarly. Define $h^{(i)}: V \rightarrow [0, 1]$ by $h^{(i)} := 1_{\mathfrak{B}''^{-1}(1, \mathfrak{B}_1(i), s)}$.

We compute p_2 as

$$\begin{aligned} p_2 &= \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in \mathbb{F}_p^{n_1}} g^{(i)}(L_i(\mathbf{x})) h^{(i)}(L_i(\mathbf{x})) \right] \\ &= \mathbb{E}_{\mathbf{x}} \left[\prod_i \left(g_{\text{str}}^{(i)}(L_i(\mathbf{x})) + g_{\text{sml}}^{(i)}(L_i(\mathbf{x})) + g_{\text{psr}}^{(i)}(L_i(\mathbf{x})) \right) h^{(i)}(L_i(\mathbf{x})) \right] \\ &\geq \mathbb{E}_{\mathbf{x}} \left[\prod_i \left(g_{\text{str}}^{(i)}(L_i(\mathbf{x})) + g_{\text{sml}}^{(i)}(L_i(\mathbf{x})) \right) h^{(i)}(L_i(\mathbf{x})) \right] - 3^{p^{n_1}} \eta(\|\mathfrak{B}''\|). \end{aligned}$$

The inequality follows from Lemma 4.4(ii) and the counting lemma, (3.1).

Write p_3 for the expectation in the last line above. We have $p_2 \geq p_3 - 3^{p^{n_1}} \eta(\|\mathfrak{B}''\|)$. Expanding the product, there are at most $2^{p^{n_1}}$ terms involving $g_{\text{sml}}^{(\mathbf{j})}$ for some $\mathbf{j} \in \mathbb{F}_p^{n_1}$. Each of these is bounded in magnitude by

$$\mathbb{E}_{\mathbf{x}} \left[\left| g_{\text{sml}}^{(\mathbf{j})}(L_{\mathbf{j}}(\mathbf{x})) \right| \prod_i h^{(i)}(L_i(\mathbf{x})) \right].$$

By applying a change of coordinates, we can transform to the case that $\mathbf{j} = 0$ (and $L_0(\mathbf{x}) = x_0$). Then by the Cauchy-Schwarz inequality, the square of the above expression is bounded by

$$\mathbb{E}_{x_0} \left[\left| g_{\text{sml}}^{(0)}(x_0) \right|^2 h^{(0)}(x_0) \right] \mathbb{E}_{x_0} \left[h^{(0)}(x_0) \mathbb{E}_{x_1, \dots, x_{n_1}} \left[\prod_{i \neq 0} h^{(i)}(L_i(\mathbf{x})) \right]^2 \right].$$

The first term is at most $\theta(\|\mathfrak{B}\|^2 \|h^{(0)}\|_2^2)$ by the fact that $(1, \mathfrak{B}_1(i), s)$ is a regular atom for all i . The second term can be counted by equidistribution applied to the system \mathbf{L}' of $2p^{n_1} - 1$ linear forms in $2n_1 - 1$ variables defined as follows. Set

$$L'_0(x_0, x_1, \dots, x_{n_1}, x'_1, \dots, x'_{n_1}) := x_0,$$

and for $\mathbf{i} \in \mathbb{F}_p^{n_1} \setminus \{0\}$, define

$$L'_{\mathbf{i},1}(x_0, x_1, \dots, x_{n_1}, x'_1, \dots, x'_{n_1}) := x_0 + i_1 x_1 + \dots + i_{n_1} x_{n_1},$$

$$L'_{\mathbf{i},2}(x_0, x_1, \dots, x_{n_1}, x'_1, \dots, x'_{n_1}) := x_0 + i_1 x'_1 + \dots + i_{n_1} x'_{n_1}.$$

By Lemma 3.23, we know that $\|\mathfrak{B}''\| \cdot |\Phi_{I''}(\mathbf{L}')| = |\Phi_{I''}(\mathbf{L})|^2$ (see also [7, Lemma 5.13]).

Thus by equidistribution, Theorem 3.19, and the rank bound on \mathfrak{B}'' , we have that the second term is at most

$$\frac{1}{|\Phi_{I''}(\mathbf{L}')|} + \alpha(\|\mathfrak{B}''\|) = \frac{\|\mathfrak{B}''\|}{|\Phi_{I''}(\mathbf{L})|^2} + \alpha(\|\mathfrak{B}''\|) \leq \frac{2\|\mathfrak{B}''\|}{|\Phi_{I''}(\mathbf{L})|^2}.$$

Applying equidistribution again we have that the first term is at most

$$\theta^2 \left(\frac{1}{\|\mathfrak{B}''\|} + \alpha(\|\mathfrak{B}''\|) \right) \leq \frac{2\theta^2}{\|\mathfrak{B}''\|}.$$

Combining these bounds and summing over all terms that contain some $g_{\text{sml}}^{(j)}$, we see that

$$p_3 \geq \mathbb{E}_{\mathbf{x}} \left[\prod_{\mathbf{i}} g_{\text{str}}^{(\mathbf{i})}(L_{\mathbf{i}}(\mathbf{x})) h^{(\mathbf{i})}(L_{\mathbf{i}}(\mathbf{x})) \right] - 2^{p^{n_1}+1} \frac{\theta}{|\Phi_{I''}(\mathbf{L})|}.$$

Write p_4 for the expectation in the last line. The quantity $g_{\text{str}}^{(\mathbf{i})}(L_{\mathbf{i}}(\mathbf{x}))$ is the density of $\chi(\mathfrak{B}_1(\mathbf{i}))$ in the atom of \mathfrak{B}'' that $L_{\mathbf{i}}(\mathbf{x})$ lies in. When $\mathfrak{B}(L_{\mathbf{i}}(\mathbf{x})) = (1, \mathfrak{B}_1(\mathbf{i}), s)$, the choice of χ implies that this density is at least $1/|\mathcal{S}|$. Thus

$$p_4 \geq \frac{1}{|\mathcal{S}|^{p^{n_1}}} \mathbb{E}_{\mathbf{x}} \left[\prod_{\mathbf{i}} h^{(\mathbf{i})}(L_{\mathbf{i}}(\mathbf{x})) \right].$$

Write p_5 for the expectation in the last line. Unwrapping the definition of $h^{(\mathbf{i})}$, this can be written as

$$p_5 = \Pr_{\mathbf{x}} (\mathfrak{B}''(L_{\mathbf{i}}(\mathbf{x})) = (1, \mathfrak{B}_1(\mathbf{i}), s) \text{ for all } \mathbf{i} \in \mathbb{F}_p^{n_1}).$$

We claim that $((1, \mathfrak{B}_1(\mathbf{i}), s))_{\mathbf{i} \in \mathbb{F}_p^{n_1}}$ is an \mathbf{L} -consistent tuple of atoms. We check this coordinate by coordinate. Obviously $(\mathfrak{B}_1(\mathbf{i}))_{\mathbf{i} \in \mathbb{F}_p^{n_1}}$ is \mathbf{L} -consistent. Furthermore any constant tuple is also obviously \mathbf{L} -consistent (this follows from the fact that \mathbf{L} is a translation-invariant pattern). Thus by another application of equidistribution and the rank bound on \mathfrak{B}'' , we have

$$p_5 \geq \frac{1}{|\Phi_{I''}(\mathbf{L}')|} - \alpha(\|\mathfrak{B}''\|) \geq \frac{1}{2|\Phi_{I''}(\mathbf{L}')|}.$$

Combining all these inequalities, we see that

$$p_1 \geq \frac{1}{|\mathcal{S}|^{p^{n_1}}} \frac{1}{2|\Phi_{I''}(\mathbf{L}')|} - 2^{p^{n_1}+1} \frac{\theta}{|\Phi_{I''}(\mathbf{L})|} - 3^{p^{n_1}} \eta(\|\mathfrak{B}''\|) - \frac{p^{n_1+1}}{|V|}.$$

This expression is positive by the definition of θ, η and the assumption that $\dim V \geq n_{\text{ramsey}}'$.

Thus we have defined a function $\chi: A_I \rightarrow \mathcal{S}$ and found linearly independent $x_0, x_1, \dots, x_{n_1} \in V$ with several desirable properties. Define $z := x_0$ and $U := \text{span}\{x_1, \dots, x_{n_1}\}$. Since $P: V \rightarrow \mathbb{F}_p$ is a linear function and $P(x_0 + i_1 x_1 + \dots + i_{n_1} x_{n_1}) = 1$ for all $\mathbf{i} \in \mathbb{F}_p^{n_1}$, we conclude that $P(z) = P(x_0) = 1$ and U is contained in the zero set of P . Furthermore, $\mathfrak{B}(x_0 + i_1 x_1 + \dots + i_{n_1} x_{n_1}) = \mathfrak{B}_1(i_1, \dots, i_{n_1})$ for all $\mathbf{i} \in \mathbb{F}_p^{n_1}$. Since \mathfrak{B}_1 was chosen such that $\text{rank } \mathfrak{B}_1 \geq r_0 + p$ and $n_1 \geq n_0$, we see that $\dim U \geq n_0$ and $\text{rank } \mathfrak{B}|_{\text{span}\{x_0, U\}} \geq r_0 + p$. By Lemma 3.16, we conclude that $\text{rank } \mathfrak{B}|_U \geq r_0$. Finally, $f(x_0 + i_1 x_1 + \dots + i_{n_1} x_{n_1}) = \chi(\mathfrak{B}_1(i_1, \dots, i_{n_1})) = \chi(\mathfrak{B}(x_0 + i_1 x_1 + \dots + i_{n_1} x_{n_1}))$ for all $\mathbf{i} \in \mathbb{F}_p^{n_1}$, which proves the desired result. \square

5.4. Main patching result. We boost the Ramsey dichotomy result to a supersaturation dichotomy result which is our main patching theorem. For our application, we will need this result to hold inside a “subvariety” of a vector space, i.e., the zero set of a sufficiently high-rank collection of nonclassical polynomials. Also for technical reasons, this supersaturation argument works only for full dimensional patterns (recall Definition 3.21).

Despite the complicated nature of this statement, the high-level overview of the proof is quite simple. If we are in case (b) of Theorem 5.6, we know that there exists an \mathcal{H} -instance not just in V , but in essentially every n_{dich} -dimensional subspace of V . Adding up over all subspaces of this dimension gives us a lot of \mathcal{H} -instances.

There are two issues with this proof sketch which we have to fix. First, it is not true that for every subspace U , the polynomial factor $\mathfrak{B}|_U$ has high rank. Fortunately, we will be able to show that a positive proportion of these subspaces have sufficiently high rank. Second, we have overcounted some of the \mathcal{H} -instances since each instance lives in many n_{dich} -dimensional subspace. To deal with this issue, we will have to show that the \mathcal{H} -instances are “distributed evenly” in some sense. The most elegant way which we are aware of for carrying out this argument is through a Cauchy–Schwarz trick, which is (5.1) below. It is this last argument which we are only able to carry out for full dimensional patterns.

THEOREM 5.9 (patching). *Fix a prime p , a finite set \mathcal{S} with an \mathbb{F}_p^\times -action, parameter lists $I, I' \in \mathcal{I}_p$ satisfying $I \leq I'$, and a positive integer ℓ_0 . There exist constants $n_{\text{patch}} = n_{\text{patch}}(p, |\mathcal{S}|, I', \ell_0)$ and $\beta_{\text{patch}} = \beta_{\text{patch}}(p, |\mathcal{S}|, I, \ell_0) > 0$ and a nondecreasing function $r_{\text{patch}} = r_{\text{patch}}(p, |\mathcal{S}|, I, \ell_0): \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ such that the following holds. Let \mathcal{H} be a finite set of \mathcal{S} -colored, I -labeled patterns such that each pattern is defined by a full dimension system of linear forms in at most ℓ_0 variables (recall Definition 3.21). Either*

- (a) *there exists a projective $\xi: \mathbb{F}_p \times A_I \rightarrow \mathcal{S}$ that does not canonically induce \mathcal{H} , or*
- (b) *for every finite-dimensional \mathbb{F}_p -vector space V satisfying $\dim V \geq n_{\text{patch}}$, every projective function $f: V \rightarrow \mathcal{S}$, every polynomial factor \mathfrak{B} on V with parameters I that satisfies $\text{rank } \mathfrak{B} \geq r_{\text{patch}}(\deg \mathfrak{B}, \|\mathfrak{B}\|)$, and every polynomial factor \mathfrak{B}' on V with parameters I' that refines \mathfrak{B} and satisfies $\text{rank } \mathfrak{B}' \geq r_{\text{patch}}(\deg \mathfrak{B}', \|\mathfrak{B}'\|)$, there is a pattern $H \in \mathcal{H}$ such that in (f, \mathfrak{B}) , the relative density of H in $\mathfrak{B}'^{-1}(A_I \times \{0\})$ is at least β_{patch} .*

Note that n_{patch} may depend on I' , but critically β_{patch} does not depend on I' .

Proof. First we define several parameters.

Write $r := r_{\text{dich}}(p, |\mathcal{S}|, I, \ell_0)$ for brevity. Define

$$n_1 := \max\{n_{\text{dich}}(p, |\mathcal{S}|, I, \ell_0), n_{\text{high-rank}}(p, I, r)\},$$

where n_0 is defined in Lemma 3.15.

Define the constants

$$n_{\text{patch}}(p, |\mathcal{S}|, I', \ell_0) := 2p^{n_1} \lceil \log_p(\|I'\|) \rceil$$

and

$$\beta_{\text{patch}}(p, |\mathcal{S}|, I, \ell_0) := \frac{1}{6400\ell_0^2 p^{2n_1 \cdot \ell} \|I\|^{2p^{n_1}}}$$

and define the nonincreasing function $\alpha: \mathbb{Z}_{>0} \rightarrow (0, 1)$ by

$$\alpha(N) := \frac{1}{2N^{2p^{n_1}}}$$

and the nondecreasing function $r_{\text{patch}}(p, |\mathcal{S}|, I, \ell_0): \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ by

$$r_{\text{patch}}(p, |\mathcal{S}|, I, \ell_0)(D, N) := r_{\text{equi}}(p, D, \alpha(N)).$$

We now proceed to the proof. Let \mathcal{H} be a finite set of \mathcal{S} -colored, I -labeled patterns each defined by a full dimension system of linear forms in at most ℓ_0 variables. We apply Theorem 5.6 to \mathcal{H} . If Theorem 5.6(a) holds, then clearly conclusion (a) holds. Now assume that Theorem 5.6(b) holds. We wish to show conclusion (b).

Let V be a finite-dimensional \mathbb{F}_p -vector space satisfying $\dim V \geq n_{\text{patch}}$, let $f: V \rightarrow \mathcal{S}$ be a projective function, let \mathfrak{B} be a polynomial factor on V with parameters I that satisfies $\text{rank } \mathfrak{B} \geq r_{\text{patch}}(\|\mathfrak{B}\|)$, and let \mathfrak{B}' be a polynomial factor on V with parameters I' that refines \mathfrak{B} and satisfies $\text{rank } \mathfrak{B}' \geq r_{\text{patch}}(\|\mathfrak{B}'\|)$.

Write $X := \mathfrak{B}'^{-1}(A_I \times \{0\})$ and $n := \dim V$. By assumption, $n \geq n_{\text{patch}}$. We wish to count \mathcal{H} instances in (f, \mathfrak{B}) that are contained in X . Write $\mathcal{H} = \bigsqcup_{\ell=1}^{\ell_0} \mathcal{H}_\ell$ where \mathcal{H}_ℓ is defined to be the subset of \mathcal{H} consisting of colored labeled patterns defined by a system of linear forms in exactly ℓ variables. We define sets $\mathfrak{U}_1, \dots, \mathfrak{U}_{\ell_0}$ as follows. \mathfrak{U}_ℓ is the set of ℓ -dimensional subspaces U of V which satisfy the following:

- $U \subseteq X$;
- there exists a colored labeled pattern $H \in \mathcal{H}_\ell$ in ℓ variables and a generic H -instance $x_1, \dots, x_\ell \in U$.

Note that the requirement that $x_1, \dots, x_\ell \in U$ are generic implies that $U = \text{span}\{x_1, \dots, x_\ell\}$. Thus $\sum_{\ell=1}^{\ell_0} |\mathfrak{U}_\ell|$ is a lower bound on the number of \mathcal{H} -instances in X .

Define the following counting function $c: \mathfrak{U}_\ell \rightarrow \mathbb{Z}_{\geq 0}$ for each $\ell \in [\ell_0]$ as follows. For $U \in \mathfrak{U}_\ell$, let $c(U)$ be the number of n_1 -dimensional subspaces that contain U and are contained in X . An application of the Cauchy-Schwarz inequality implies that

$$(5.1) \quad |\mathfrak{U}_\ell| \geq \frac{(\sum_{U \in \mathfrak{U}_\ell} c(U))^2}{\sum_{U \in \mathfrak{U}_\ell} c(U)^2}.$$

Define S_1 to be the number of n_1 -dimensional subspaces W of V such that $W \subseteq X$ and $\text{rank}(\mathfrak{B}|_W) \geq r$. By Theorem 5.6(b), every such W contains a generic \mathcal{H} -instance. Thus

$$\sum_{\ell=1}^{\ell_0} \sum_{U \in \mathfrak{U}_\ell} c(U) \geq S_1.$$

By the pigeonhole principle, there exists some $\ell \in [\ell_0]$ such that

$$(5.2) \quad \sum_{U \in \mathfrak{U}_\ell} c(U) \geq \frac{S_1}{\ell_0}.$$

We fix such a value of $\ell \in [\ell_0]$ for the rest of the proof.

Define S_2 to be the number of ordered n_1 -tuples $(x_1, \dots, x_{n_1}) \in V^{n_1}$ which satisfy the following:

- x_1, \dots, x_{n_1} are linearly independent;
- $\text{span}\{x_1, \dots, x_{n_1}\} \subseteq X$;
- $\text{rank}(\mathfrak{B}|_{\text{span}\{x_1, \dots, x_{n_1}\}}) \geq r$.

We can compute

$$(5.3) \quad S_2 = S_1 \prod_{i=0}^{n_1-1} (p^{n_1} - p^i) \leq p^{n_1^2} S_1.$$

Define S_3 to be the number of ordered n_1 -tuples $(x_1, \dots, x_{n_1}) \in V^{n_1}$ such that $\text{span}\{x_1, \dots, x_{n_1}\} \subseteq X$ and $\text{rank}(\mathfrak{B}|_{\text{span}\{x_1, \dots, x_{n_1}\}}) \geq r$. We can easily bound

$$(5.4) \quad S_2 \geq S_3 - p^{n \cdot n_1} p^{n_1 - n}.$$

By the definition of n_1 and Lemma 3.15, there exists a polynomial factor \mathfrak{B}_1 on $\mathbb{F}_p^{n_1}$ with parameters I and rank at least r . Define S_4 to be the number of ordered n_1 -tuples $(x_1, \dots, x_{n_1}) \in V^{n_1}$ such that $\mathfrak{B}_1(i_1, \dots, i_{n_1}) = \mathfrak{B}(i_1 x_1 + \dots + i_{n_1} x_{n_1})$ and $\text{span}\{x_1, \dots, x_{n_1}\} \subseteq X$. Notice that $S_3 \geq S_4$.

Write $\mathbf{L}' := \mathbf{L}^{n_1}$, the system of p^{n_1} linear forms in n_1 variables that define an n_1 -dimensional subspace (see Definition 2.7). By definition, $(P_{d,k}^i(i_1, \dots, i_{n_1}))_{i \in \mathbb{F}_p^{n_1}} \in \Phi_{d,k}(\mathbf{L}')$ for every $(d, k) \in D_p$ and $i \in [I_{d,k}]$ where $P_{d,k}^i$ is the i th nonclassical polynomial of degree d and depth k defining \mathfrak{B}_1 . Also, $(0, \dots, 0) \in \Phi_{d,k}(\mathbf{L}')$ for all $(d, k) \in D_p$. Define $\mathbf{a} \in A_{I'}^{\mathbb{F}_p^{n_1}}$ by

$$(a_i)_{d,k}^i = \begin{cases} P_{d,k}^i(i_1, \dots, i_{n_1}) & \text{if } i \leq I_{d,k}, \\ 0 & \text{if } I_{d,k} < i \leq I'_{d,k}. \end{cases}$$

By the above discussion, \mathbf{a} is \mathbf{L}' -consistent, so by Theorem 3.19 and the rank assumption of \mathfrak{B}' , we find

$$(5.5) \quad \begin{aligned} S_3 \geq S_4 &= |\{\mathbf{x} \in V^{n_1} : \mathfrak{B}'(L'_i(\mathbf{x})) = a_i \text{ for all } i \in \mathbb{F}_p^{n_1}\}| \\ &\geq \left(\frac{1}{|\Phi_{I'}(\mathbf{L}')|} - \alpha(\|I'\|) \right) p^{n \cdot n_1} \geq \frac{p^{n \cdot n_1}}{2|\Phi_{I'}(\mathbf{L}')|}. \end{aligned}$$

Combining (5.2), (5.3), (5.4), and (5.5), we conclude

$$(5.6) \quad \sum_{U \in \mathcal{U}_\ell} c(U) \geq \frac{1}{\ell_0 p^{n_1^2}} p^{n \cdot n_1} \left(\frac{1}{2|\Phi_{I'}(\mathbf{L}')|} - p^{n_1 - n} \right) \geq \frac{1}{4\ell_0 p^{n_1^2} |\Phi_{I'}(\mathbf{L}')|} p^{n \cdot n_1}.$$

Next we find an upper bound on $\sum_{U \in \mathcal{U}_\ell} c(U)^2$. Define T_1 to be the number of triples (U, W, W') where U is an ℓ -dimensional subspace of V and W, W' are both n_1 -dimensional subspaces of V that contain U and are contained in X . First note that

$$(5.7) \quad \sum_{U \in \mathcal{U}_\ell} c(U)^2 \leq T_1.$$

Now define T_2 to be the number of ordered $(2n_1 - \ell)$ -tuples $(x_1, \dots, x_\ell, y_1, \dots, y_{n_1 - \ell}, z_1, \dots, z_{n_1 - \ell}) \in V^{2n_1 - \ell}$ such that $x_1, \dots, x_\ell, y_1, \dots, y_{n_1 - \ell}$ are linearly independent, $x_1, \dots, x_\ell, z_1, \dots, z_{n_1 - \ell}$ are linearly independent, $\text{span}\{x_1, \dots, x_\ell, y_1, \dots, y_{n_1 - \ell}\} \subseteq X$, and $\text{span}\{x_1, \dots, x_\ell, z_1, \dots, z_{n_1 - \ell}\} \subseteq X$. We compute

$$(5.8) \quad T_2 = \left(\prod_{i=0}^{\ell-1} (p^\ell - p^i) \right) \left(\prod_{i=\ell}^{n_1-1} (p^{n_1} - p^i) \right)^2 T_1 \geq \frac{p^{2n_1(n_1 - \ell)}}{100} T_1.$$

Next define T_3 to be the number of ordered $(2n_1 - \ell)$ -tuples $(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) \in V^{2n_1-\ell}$ such that $\text{span}\{x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}\} \subseteq X$ and $\text{span}\{x_1, \dots, x_\ell, z_1, \dots, z_{n_1-\ell}\} \subseteq X$. Clearly $T_2 \leq T_3$.

Define \mathbf{L}'' to be the following system of $2p^{n_1} - p^\ell$ linear forms in $2n_1 - n_\ell$ variables. For $\mathbf{i} \in \mathbb{F}_p^\ell$, define

$$L''_{\mathbf{i}}(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) := i_1 x_1 + \dots + i_\ell x_\ell.$$

For $\mathbf{i} \in \mathbb{F}_p^{n_1} \setminus (\mathbb{F}_p^\ell \times \{0\}^{n_1-\ell})$, define

$$\begin{aligned} L''_{i,1}(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) &:= i_1 x_1 + \dots + i_\ell x_\ell + i_{\ell+1} y_1 + \dots + i_{n_1} y_{n_1-\ell}, \\ L''_{i,2}(x_1, \dots, x_\ell, y_1, \dots, y_{n_1-\ell}, z_1, \dots, z_{n_1-\ell}) &:= i_1 x_1 + \dots + i_\ell x_k + i_{\ell+1} z_1 + \dots + i_{n_1} z_{n_1-\ell}. \end{aligned}$$

Now let \mathfrak{B}'' be the polynomial factor on V with parameters $I' - I$ defined by homogeneous nonclassical polynomials

$$(P_{d,k}^i)^{(d,k) \in D_p}_{I_{d,k} < i \leq I'_{d,k}},$$

where $P_{d,k}^i$ is the i th nonclassical polynomial of degree d and depth k that defines \mathfrak{B}' . The important property of this polynomial factor is that $X = \mathfrak{B}''^{-1}(\mathbf{0})$. Also note that $\text{rank } \mathfrak{B}'' \geq \text{rank } \mathfrak{B}' \geq r(\|\mathfrak{B}'\|)$. By Theorem 3.19 and the bounds on $\text{rank } \mathfrak{B}''$, we find

$$\begin{aligned} T_2 \leq T_3 &= |\{\mathbf{x} \in V^{2n_1-\ell} : \mathfrak{B}''(\mathbf{L}''(\mathbf{x})) = \mathbf{0}\}| \\ (5.9) \quad &\leq \left(\frac{1}{|\Phi_{I'-I}(\mathbf{L}'')|} + \alpha(\|I'\|) \right) p^{n(2n_1-\ell)} \leq \frac{2p^{n(2n_1-\ell)}}{|\Phi_{I'-I}(\mathbf{L}'')|}. \end{aligned}$$

Combining (5.7), (5.8), and (5.9), we conclude

$$(5.10) \quad \sum_{U \in \mathfrak{U}_\ell} c(U)^2 \leq \frac{200p^{2n_1-\ell}}{p^{2n_1^2} |\Phi_{I'-I}(\mathbf{L}'')|} p^{n(2n_1-\ell)}.$$

Finally, we combine (5.1), (5.6), and (5.10) to find

$$|\mathfrak{U}_\ell| \geq \frac{|\Phi_{I'-I}(\mathbf{L}'')|}{3200\ell_0^2 p^{2n_1-\ell} |\Phi_{I'}(\mathbf{L}')|^2} p^{n \cdot \ell}.$$

Consider \mathbf{L}^ℓ , the system of p^ℓ linear forms in ℓ variables that define an ℓ -dimensional subspace (see Definition 2.7). By Lemma 3.23, we have $|\Phi_{I'-I}(\mathbf{L}^\ell)| \cdot |\Phi_{I'-I}(\mathbf{L}'')| = |\Phi_{I'-I}(\mathbf{L}')|^2$. Thus the above expression simplifies to

$$|\mathfrak{U}_\ell| \geq \frac{1}{3200\ell_0^2 p^{2n_1-\ell} |\Phi_{I'}(\mathbf{L}')|^2} \cdot \frac{p^{n \cdot \ell}}{|\Phi_{I'-I}(\mathbf{L}^\ell)|} \geq \frac{1}{3200\ell_0^2 p^{2n_1-\ell} \|I\|^{2p^{n_1}}} \cdot \frac{p^{n \cdot \ell}}{|\Phi_{I'-I}(\mathbf{L}^\ell)|}.$$

Therefore there exists some colored labeled pattern $H = (\mathbf{L}, \psi, \phi) \in \mathcal{H}_\ell$ where \mathbf{L} is a full dimension system of linear forms in ℓ variables and such that the number of generic H -instances in (f, \mathfrak{B}) which are contained in X is at least $1/|\mathcal{H}_\ell|$ times the right-hand side of the above equation. Note that by equidistribution, Theorem 3.19, and the rank bound on \mathfrak{B}'' ,

$$(5.11) \quad \Lambda_{\mathbf{L}}(1_X, \dots, 1_X) \leq \frac{1}{|\Phi_{I'-I}(\mathbf{L})|} + \alpha(\|I'\|) \leq \frac{2}{|\Phi_{I'-I}(\mathbf{L})|}.$$

Noting that since \mathbf{L} is full dimensional, we have $|\Phi_{I'-I}(\mathbf{L})| = |\Phi_{I'-I}(\mathbf{L}^\ell)|$. Thus dividing the two above quantities, we find that the relative density of the above H in X is at least

$$\frac{1}{6400\ell_0^2 p^{2n_1 \cdot \ell} \|I\|^{2p^{n_1}}} = \beta_{\text{patch}}(p, |\mathcal{S}|, I, \ell_0).$$

□

6. Proof of removal lemmas. In this section we prove the main technical result, Theorem 2.8, the projective removal lemma. Then we give a short deduction of the full removal lemma, Theorem 2.4, from it.

Notation. As usual, for an atom $a \in A_I$ (defined in (3.3)), we use $a_{d,k} \in \mathbb{U}_{k+1}^{I_{d,k}}$ to denote the degree d , depth k part of a . We use the notation $\tilde{A}_I \subset A_I$ to denote the set $\tilde{A}_I := \{a \in A_I : a_{1,0} = 0\}$. This is the set of atoms that are not regularized by Theorem 4.6. Also define $\tilde{I} \in \mathcal{I}_p$ by $\tilde{I}_{1,0} = 0$ and $\tilde{I}_{d,k} = I_{d,k}$ otherwise.

Overview of proof. There are four main steps. First, we apply a “compactness argument” inspired by a similar argument of Alon and Shapira. We will talk about this more later, but for now let us say that this essentially allows us to assume that \mathcal{H} is a finite set of colored patterns.

The rest of the proof involves “cleaning up” the coloring $f: V \rightarrow \mathcal{S}$ so that every pattern $H \in \mathcal{H}$ appears in f either zero or many times. The next step is to apply our subatom selection theorem. This produces a polynomial factor \mathfrak{B} , a refinement \mathfrak{B}' , and a subatom selection function $s: A_I \rightarrow A_{I'}$. Define X to be the “subvariety” of V that is the union of the subatoms selected by s . We are guaranteed that X is a “regular model” for V in the sense that most subatoms of X are regular and are colored similarly to the corresponding atom they are contained in.

The third step is to “clean up” the regular atoms. For each $a \in A_I \setminus \tilde{A}_I$, we recolor any point colored with a color that is low-density in the subatom $s(a)$. After this cleaning procedure, arithmetic regularity implies that any pattern which appears in the regular atoms of the cleaned coloring must appear a lot (namely, with positive density) in the regular model X in the original coloring.

To complete the proof we apply patching to handle the irregular atoms. By our patching result, there exists some ξ such that every H that ξ canonically induces appears a lot in the irregular atoms of X . By replacing our coloring on the irregular atoms with the ξ -canonical coloring, we produce a new coloring. The property of the new coloring is that any H which appears in the new coloring must also have appeared with positive density in the regular model X in the original coloring, which is sufficient to prove the removal lemma.

As in the proof of Lemma 5.8, there is quite a bit of technical work involved with the application of the arithmetic regularity lemma, but the above contains the high-level ideas of the proof.

We now give the definitions necessary to run the compactness argument. The recoloring produced in the above argument can be summarized by two pieces of data. First, we record which colors appear in each regular atom, and second, we record which ξ -canonical coloring was used on the irregular atoms. With these two pieces of data we can determine which patterns we could possibly find in the recoloring.

DEFINITION 6.1. Fix a prime p , a finite set \mathcal{S} equipped with an \mathbb{F}_p^\times -action, and a parameter list $I \in \mathcal{I}_p$. A summary function with parameters I is a pair (F, ξ) consisting of a function $F: (A_I \setminus \tilde{A}_I) \rightarrow 2^{\mathcal{S}} \setminus \{\emptyset\}$ and a projective function $\xi: \mathbb{F}_p \times A_{\tilde{I}} \rightarrow \mathcal{S}$.

DEFINITION 6.2. For an \mathcal{S} -colored pattern $H = (\mathbf{L}, \psi)$ consisting of m linear forms and a summary function (F, ξ) with parameters I , say that (F, ξ) partially induces H if there exists a tuple of atoms $\mathbf{a} \in A_I^m$ such that the following holds:

- (i) \mathbf{a} is \mathbf{L} -consistent, i.e., $\mathbf{a} \in \Phi_I(\mathbf{L})$;
- (ii) for each $i \in [m]$ such that $a_i \notin \tilde{A}_I$, we have $\psi(i) \in F(a_i)$;
- (iii) defining $J := \{i \in [m] : a_i \in \tilde{A}_I\}$ and $H_J := ((L_i)_{i \in J}, \psi|_J, (a_i)_{i \in J})$, an \mathcal{S} -colored I -labeled pattern, we have ξ canonically induces H_J .

Proof of Theorem 2.8. We are given a parameter $\epsilon > 0$ and a possibly infinite set \mathcal{H} of \mathcal{S} -colored patterns over \mathbb{F}_p of the form $(\bar{\mathbf{L}}^\ell, \psi)$ where ℓ is some positive integer and $\psi: E_\ell \rightarrow \mathcal{S}$ is some map. (See Definition 2.7 for the definition of $\bar{\mathbf{L}}^\ell$ and E_ℓ .)

We begin with a “compactness argument” based on ideas of Alon and Shapira that allows us to reduce to the case when \mathcal{H} is finite size.

For each parameter list $I \in \mathcal{I}_p$, we define a finite subset $\mathcal{H}_I \subseteq \mathcal{H}$ as follows. Consider the set of all summary functions (F, ξ) with parameters I . If there exists any $H \in \mathcal{H}$ such that (F, ξ) partially induces H , include one such H in \mathcal{H}_I . Note that $|\mathcal{H}_I|$ is at most the number of summary functions with parameters I , which is finite.

Define the compactness functions $\Psi_{\mathcal{H}}: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ as follows. Let $\Psi_{\mathcal{H}}(D, N)$ be the largest positive integer ℓ such that there exists a parameter list $I \in \mathcal{I}_p$ satisfying $\deg I \leq D$ and $\|I\| \leq N$ such that a pattern of the form $(\bar{\mathbf{L}}^\ell, \psi)$ exists in \mathcal{H}_I .

Now we set several parameters. Define nonincreasing functions $\eta: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ by

$$\eta(D, N) := \frac{1}{40} \left(\frac{\epsilon}{12N|\mathcal{S}|} \right)^{p^{\Psi_{\mathcal{H}}(D, N)}},$$

$\beta: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ by

$$\beta(D, N) := \min_{I \in \mathcal{I}_p: \deg I \leq D, \|I\| \leq N} \beta_{\text{patch}}(p, |\mathcal{S}|, I, \Psi_{\mathcal{H}}(D, N)),$$

$\theta: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ by

$$\theta(D, N) := \frac{\beta(D, N)}{40} \left(\frac{\epsilon}{8|\mathcal{S}|} \right)^{p^{\Psi_{\mathcal{H}}(D, N)}},$$

and $\alpha: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow (0, 1)$ by

$$\alpha(D, N) := \frac{\beta(D, N)}{10N2^{p^{\Psi_{\mathcal{H}}(D, N)}}}.$$

Then define nondecreasing functions $r: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ by

$$r(D, N) := \max\{r_{\text{equi}}(p, D, \alpha(D, N)), \max_{I \in \mathcal{I}_p: \deg I \leq D, \|I\| \leq N} r_{\text{patch}}(p, |\mathcal{S}|, I, \Psi_{\mathcal{H}}(D, N))(D, N) + p \lceil \log_p N \rceil\}$$

and $d: \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \rightarrow \mathbb{Z}_{>0}$ by

$$d(D, N) := p^{\Psi_{\mathcal{H}}(D, N)}.$$

Define parameters

$$\zeta := \frac{\epsilon}{16|\mathcal{S}|} \quad \text{and} \quad c_0 := \lceil \log_p(2/\epsilon) \rceil.$$

Then define

$$C_{\max} := C_{\text{reg}}(p, |\mathcal{S}|, c_0, \zeta, \eta, \theta, d, r),$$

$$D_{\max} := D_{\text{reg}}(p, |\mathcal{S}|, c_0, \zeta, \eta, \theta, d, r),$$

$$n_{\min} := \max \left\{ n_{\text{reg}}(p, c_0, \zeta), \max_{I \in \mathcal{I}_p: \deg I \leq D_{\max}, \|I\| \leq C_{\max}} n_{\text{patch}}(p, |\mathcal{S}|, I, \Psi_{\mathcal{H}}(D_{\max}, C_{\max})) + \lceil \log_p C_{\max} \rceil \right\}.$$

Finally, define

$$\delta(\epsilon, \mathcal{H}) := \min \left\{ \frac{\beta(D_{\max}, C_{\max})}{40} \left(\frac{\epsilon}{4C_{\max}|\mathcal{S}|} \right)^{p^{\Psi_{\mathcal{H}}(D_{\max}, C_{\max})}}, p^{-n_{\min} \cdot \Psi_{\mathcal{H}}(D_{\max}, C_{\max})} \right\}$$

and

$$\mathcal{H}_{\epsilon} := \bigcup_{I \in \mathcal{I}_p: \|I\| \leq C_{\max}, \deg I \leq D_{\max}} \mathcal{H}_I.$$

Since the union is over a finite set of I , we have \mathcal{H}_{ϵ} is finite. We will show that this choice of δ and \mathcal{H}_{ϵ} satisfies the desired conclusion.

Let V be a finite-dimensional \mathbb{F}_p -vector space and $f: V \rightarrow \mathcal{S}$ be a projective function with H -density at most $\delta(\epsilon, \mathcal{H})$ for every $H \in \mathcal{H}_{\epsilon}$. Our goal is to produce a projective recoloring $g: V \rightarrow \mathcal{S}$ that agrees with f on all but an at most ϵ -fraction of V that has no generic H -instances for every $H \in \mathcal{H}$.

First note that if $\dim V < n_{\min}$, the theorem easily follows. This is because for the pattern $H = (\bar{\mathbf{L}}^{\ell}, \psi)$, if there exists an H -instance in f , then the H -density in f is at least $1/|V|^{\ell}$. Thus taking $g = f$ and noticing that we chose $\delta(\epsilon, \mathcal{H}) \leq p^{-n_{\min} \cdot \Psi_{\mathcal{H}}(C_{\max}, D_{\max})}$, the theorem holds in this case.

Now assume that $\dim V \geq n_{\min}$. We apply Theorem 4.6 to the functions $\{1_{f^{-1}(c)}\}_{c \in \mathcal{S}}$ with parameters $p, |\mathcal{S}|, c_0, \zeta, \eta, \theta, d, r$. This produces a polynomial factor \mathfrak{B} and a refinement \mathfrak{B}' both on V with parameters I and I' and a subatom selection function $s: A_I \rightarrow A_{I'}$ satisfying several other desirable properties.

As above, define $\tilde{A}_I \subset A_I$ to be the set $\tilde{A}_I := \{a \in A_I : a_{1,0} = 0\}$. We call the atoms $a \in \tilde{A}_I$ *irregular* and the remaining atoms $a \in A_I \setminus \tilde{A}_I$ *regular*. Define \tilde{V} to be the codimension- $I_{1,0}$ subspace of V that is the common zero set of all $I_{1,0}$ linear polynomials defining \mathfrak{B} . The irregular atoms of \mathfrak{B} exactly consist of \tilde{V} , i.e., $\mathfrak{B}^{-1}(\tilde{A}_I) = \tilde{V}$.

Define $\tilde{\mathfrak{B}}$ to be the polynomial factor on \tilde{V} defined by the restrictions of the homogeneous nonclassical polynomials that define \mathfrak{B} to \tilde{V} , except for the linear polynomials (which restrict to the zero function). Let $\tilde{I} \in \mathcal{I}_p$ be the parameter list of $\tilde{\mathfrak{B}}$ ($\tilde{I}_{1,0} = 0$ and $\tilde{I}_{d,k} = I_{d,k}$ otherwise). Also define $\tilde{\mathfrak{B}}'$ to be the polynomial factor on \tilde{V} defined by the restrictions of the homogeneous nonclassical polynomial that define \mathfrak{B}' to \tilde{V} , except for the linear polynomials that also define \mathfrak{B} . Let $\tilde{I}' \in \mathcal{I}_p$ be the parameter list of $\tilde{\mathfrak{B}}'$ ($\tilde{I}'_{1,0} = I'_{1,0} - I_{1,0}$ and $\tilde{I}'_{d,k} = I'_{d,k}$ otherwise). Note that by Lemma 3.16 and our definition of r , we have

$$\text{rank } \tilde{\mathfrak{B}} \geq r_{\text{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg \mathfrak{B}, \|\mathfrak{B}\|))(\deg \mathfrak{B}, \|\mathfrak{B}\|),$$

$$\text{rank } \tilde{\mathfrak{B}}' \geq r_{\text{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg \mathfrak{B}, \|\mathfrak{B}\|))(\deg \mathfrak{B}', \|\mathfrak{B}'\|).$$

We will “clean up” the regular atoms by removing low-density colors in a projective manner. We will “patch” the irregular atoms by replacing the coloring by a new coloring $\Xi_{\xi, \iota, \tilde{\mathfrak{B}}}$ for some projective $\xi: \mathbb{F}_p \times A_{\tilde{I}} \rightarrow \mathcal{S}$ and $\iota: \tilde{V} \xrightarrow{\sim} \mathbb{F}_p^{\dim \tilde{V}}$.

Note that to check that the recoloring $g: V \rightarrow \mathcal{S}$ is projective, it suffices to check this fact separately on \tilde{V} and on $V \setminus \tilde{V}$.

Clean up regular atoms. For each $a \in (A_I \setminus \tilde{A}_I)$, say that a color $c \in \mathcal{S}$ is *high-density in a* if it appears in $\mathfrak{B}'^{-1}(s(a))$ with density at least $\epsilon/(4|\mathcal{S}|)$. Say that a color is *low-density in a* otherwise.

First note that a basic property of subatom selection functions, Lemma 3.25(ii), is the following. For $a \in A_I$ and $b \in \mathbb{F}_p^\times$, we have $b \cdot s(a) = s(b \cdot a)$. Combined with the projectiveness of f , this implies that for a color $c \in \mathcal{S}$ and $b \in \mathbb{F}_p^\times$, the c -density in $\mathfrak{B}'^{-1}(s(a))$ is the same as the $(b \cdot c)$ -density in $\mathfrak{B}^{-1}(s(b \cdot a))$. Thus a color c is high-density in a if and only if $b \cdot c$ is high-density in $b \cdot a$.

We pick a single high-density color $c_a \in \mathcal{S}$ for each regular atom $a \in (A_I \setminus \tilde{A}_I)$ in a projective way, namely such that $b \cdot c_a = c_{b \cdot a}$ for all $a \in (A_I \setminus \tilde{A}_I)$ and $b \in \mathbb{F}_p^\times$. By the argument in the above paragraph, this is possible.

Now we define our recoloring of the regular atoms, $g: (V \setminus \tilde{V}) \rightarrow \mathcal{S}$ as follows. For each $a \in (A_I \setminus \tilde{A}_I)$ and $x \in \mathfrak{B}^{-1}(a)$, we define $g(x) := f(x)$ unless $f(x)$ is low-density in a , in which case we define $g(x) := c_a$. Note that g is a projective function. Furthermore we claim that g differs from f on at most an $(\epsilon/2)$ -fraction of V .

By Theorem 4.6(vi), for all but at most a ζ -fraction of $a \in A_I$, the c -density in $\mathfrak{B}^{-1}(a)$ and the c -density in $\mathfrak{B}'^{-1}(s(a))$ differ by at most ζ for all $s \in \mathcal{S}$. Thus for most atoms a , each low-density color appears in $\mathfrak{B}^{-1}(a)$ with density at most $\epsilon/(4|\mathcal{S}|) + \zeta$, so f and g differ on at most an $(\epsilon/4 + \zeta|\mathcal{S}|)$ -fraction of these atoms. The functions f and g may differ completely on the other atoms, but there are at most $\zeta\|\mathfrak{B}\|$ of these. It follows by equidistribution, Theorem 3.19, and the rank bound on \mathfrak{B} that each atom of \mathfrak{B} is at most an $(\|\mathfrak{B}\|^{-1} + \alpha(\deg \mathfrak{B}, \|\mathfrak{B}\|))$ -fraction of V . Putting this all together, we see that g differs from f on at most the following fraction of V :

$$\zeta\|\mathfrak{B}\| \left(\frac{1}{\|\mathfrak{B}\|} + \alpha(\deg \mathfrak{B}, \|\mathfrak{B}\|) \right) + \left(\frac{\epsilon}{4|\mathcal{S}|} + \zeta \right) |\mathcal{S}| < \frac{\epsilon}{2}.$$

Patch irregular atoms. We define $\tilde{\mathcal{H}}$ to be the set of all \mathcal{S} -colored I -labeled patterns that are defined by a full dimension system of linear forms in at most $\Psi_{\mathcal{H}}(\deg I, \|I\|)$ variables and whose relative density in $\mathfrak{B}'^{-1}(A_{\tilde{I}} \times \{0\})$ is less than $\beta_{\text{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg I, \|I\|))$.

We apply our patching result, Theorem 5.9, to the set $\tilde{\mathcal{H}}$. Our definitions are exactly such that $f|_{\tilde{V}}$ with $\tilde{\mathfrak{B}}, \tilde{\mathfrak{B}}'$ demonstrate that Theorem 5.9(b) does not hold. In particular, we checked the rank assumptions on $\tilde{\mathfrak{B}}$ and $\tilde{\mathfrak{B}}'$ above when they were defined. Furthermore, we assumed that $\dim V \geq n_{\min}$, which implies that $\dim \tilde{V} \geq n_{\text{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg I, \|I\|))$. Finally we defined $\tilde{\mathcal{H}}$ to be the set of patterns which appear with very low density in $\mathfrak{B}'^{-1}(A_{\tilde{I}} \times \{0\})$. Thus we conclude that Theorem 5.9(a) holds. This means that there exists a projective $\xi: \mathbb{F}_p \times A_{\tilde{I}} \rightarrow \mathcal{S}$ that does not canonically induce $\tilde{\mathcal{H}}$. In particular, this means that for any fixed isomorphism $\iota: \tilde{V} \xrightarrow{\sim} \mathbb{F}_p^{\dim \tilde{V}}$, there are no generic H -instances in $(\Xi_{\xi, \iota, \tilde{\mathfrak{B}}})$ for any $H \in \tilde{\mathcal{H}}$.

We complete our definition of $g: V \rightarrow \mathcal{S}$ by defining $g(x) := \Xi_{\xi, \iota, \tilde{\mathfrak{B}}}(x)$ for all $x \in \tilde{V}$. To conclude this portion of the proof, we make sure that $|\tilde{V}| \leq (\epsilon/2)|V|$. By assumption,

$$|\tilde{V}|/|V| = p^{-I_{1,0}} \leq p^{-c_0} = \epsilon/2,$$

as desired.

Proof of correctness. We claim that g has no generic H -instances for each $H \in \mathcal{H}$. Define $F: (A_I \setminus \tilde{A}_I) \rightarrow 2^{\mathcal{S}} \setminus \{0\}$ to map a to the set of high-density colors in a and recall the projective function $\xi: \mathbb{F}_p \times A_{\tilde{I}} \rightarrow \mathcal{S}$ defined in the “patch irregular atoms” section. Now suppose that the desired conclusion does not hold, i.e., that there is a generic H' -instance in g for some $H' \in \mathcal{H}$. By the construction of g , this means that (F, ξ) partially induces H' . By the definition of the \mathcal{H}_I , this means that there is some $H \in \mathcal{H}_I \subseteq \mathcal{H}_\epsilon$ so that (F, ξ) also partially induces H . We will reach a contradiction by showing that this implies that the H -density in f is larger than $\delta(\epsilon, \mathcal{H})$.

Say that $H = (\bar{\mathbf{L}}^\ell, \psi)$ (note that $\ell \leq \Psi_{\mathcal{H}}(\deg I, \|I\|)$). Since (F, ξ) partially induces H , this implies that there exists a tuple of atoms $\mathbf{a} \in \Phi_I(\bar{\mathbf{L}}^\ell) \subseteq A_I^{E_\ell}$ with several desirable properties. Define $J := \{i \in E_\ell : a_i \in \tilde{A}_I\}$. Recalling that \tilde{A}_I is just the set of atoms whose linear part is zero, we can conclude that $J \subseteq E_\ell \subset \mathbb{F}_p^\ell$ is the intersection of E_ℓ with some linear subspace $U \leq \mathbb{F}_p^\ell$ of dimension $\ell' \leq \ell$. This means that the system $(L_i^\ell)_{i \in J}$ is equivalent to the system $(L_i^{\ell'})_{i \in J}$ where now we view $J \subset U \simeq \mathbb{F}_p^{\ell'}$. Define, $H_J := ((L_i^{\ell'})_{i \in E_\ell}, \psi|_J, (a_i)_{i \in J})$, an \mathcal{S} -colored I -labeled pattern. By Lemma 3.22, we see that H_J is a full dimension pattern.

The first property, that \mathbf{a} is $\bar{\mathbf{L}}^\ell$ -consistent, implies that $s(\mathbf{a})$ is also $\bar{\mathbf{L}}^\ell$ -consistent, by Lemma 3.25(iii). The second property implies that for each $i \in (E_\ell \setminus J)$, the color $\psi(i)$ is high-density in a_i . The third property, together with our definition of ξ implies that in (f, \mathfrak{B}) , the relative density of H_J in $\mathfrak{B}'^{-1}(A_{\tilde{I}} \times \{0\})$ is at least $\beta_{\text{patch}}(p, |\mathcal{S}|, \tilde{I}, \Psi_{\mathcal{H}}(\deg I, \|I\|) \geq \beta(\deg I, \|I\|)$.

Now we put everything together as follows. Write $f^{(i)}$ for $1_{f^{-1}(\psi(i))}$. There is a decomposition $f^{(i)} = f_{\text{str}}^{(i)} + f_{\text{sml}}^{(i)} + f_{\text{psr}}^{(i)}$ given by Theorem 4.6 for each $i \in E_\ell$. Let p_1 be the H -density in f . We lower bound p_1 as follows:

$$\begin{aligned} p_1 &= \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in E_\ell} f^{(i)}(L_i^\ell(\mathbf{x})) \right] \\ &= \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in J} f^{(i)}(L_i^\ell(\mathbf{x})) \prod_{i \in E_\ell \setminus J} \left(f_{\text{str}}^{(i)}(L_i^\ell(\mathbf{x})) + f_{\text{sml}}^{(i)}(L_i^\ell(\mathbf{x})) + f_{\text{psr}}^{(i)}(L_i^\ell(\mathbf{x})) \right) \right] \\ &\geq \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in J} f^{(i)}(L_i^\ell(\mathbf{x})) \prod_{i \in E_\ell \setminus J} \left(f_{\text{str}}^{(i)}(L_i^\ell(\mathbf{x})) + f_{\text{sml}}^{(i)}(L_i^\ell(\mathbf{x})) \right) \right] - 3^{|E_\ell|} \eta(\deg \mathfrak{B}', \|\mathfrak{B}'\|). \end{aligned}$$

The inequality follows from Theorem 4.6(iii), the counting lemma (3.1), and the fact that the complexity of H is at most $d(\deg I, \|I\|) = p^{\Psi_{\mathcal{H}}(\deg I, \|I\|)}$.

Write p_2 for the expectation in the last line above. By Theorem 4.6(iii), the expression inside the expectation is nonnegative so we can restrict the expectation to \mathbf{x} such that $\bar{\mathbf{L}}^\ell(\mathbf{x}) = s(\mathbf{a})$. Thus

$$\begin{aligned} p_2 &\geq \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in J} f^{(i)}(L_i^\ell(\mathbf{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \right. \\ &\quad \left. \prod_{i \in E_\ell \setminus J} \left(f_{\text{str}}^{(i)}(L_i^\ell(\mathbf{x})) + f_{\text{sml}}^{(i)}(L_i^\ell(\mathbf{x})) \right) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \right]. \end{aligned}$$

Write p_3 for the expectation in the last line above. Expanding the product, there are at most $2^{|E_\ell|}$ terms involving $f_{\text{sml}}^{(j)}$ for some $j \in E_\ell \setminus J$. Each of these is bounded in magnitude by

$$\mathbb{E}_{\mathbf{x}} \left[\left| f_{\text{sml}}^{(j)}(L_j(\mathbf{x})) \right| \prod_{i \in E_\ell} 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \right].$$

By applying a change of coordinates, we can assume that $L_j^\ell(\mathbf{x}) = x_1$. Then by the Cauchy–Schwarz inequality, the square of the above expression is bounded by

$$\mathbb{E}_{x_1} \left[\left| f_{\text{sml}}^{(j)}(x_1) \right|^2 1_{\mathfrak{B}'^{-1}(s(a_j))}(x_1) \right] \mathbb{E}_{x_2, \dots, x_\ell} \left[\left| \prod_{i \neq j} 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \right|^2 \right].$$

The first term is at most $\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)^2 \|1_{\mathfrak{B}'^{-1}(s(a_j))}\|_2^2$ by Theorem 4.6(vi) and the fact that a_j is a regular atom for $j \in E_\ell \setminus J$. The second term can be counted by equidistribution applied to the system \mathbf{L}' of $2|E_k| - 1$ linear forms in $2\ell - 1$ variables defined as follows. Set

$$L'_j(x_1, x_2, \dots, x_\ell, x'_2, \dots, x'_\ell) := x_1,$$

and for $i \in E_\ell \setminus \{j\}$, define

$$L'_{i,1}(x_1, x_2, \dots, x_\ell, x'_2, \dots, x'_\ell) := L_i^\ell(x_1, x_2, \dots, x_\ell),$$

$$L'_{i,2}(x_1, x_2, \dots, x_\ell, x'_2, \dots, x'_\ell) := L_i^\ell(x_1, x'_2, \dots, x'_\ell).$$

By Lemma 3.23, we know that $\|\mathfrak{B}'\| \cdot |\Phi_{\mathbf{L}'}(\mathbf{L}')| = |\Phi_{\mathbf{L}'}(\overline{\mathbf{L}}^\ell)|^2$ (see also [7, Lemma 5.13]).

Thus by equidistribution (Theorem 3.19) and the rank bound on \mathfrak{B}' , we have the second term is at most

$$\frac{1}{|\Phi_{\mathbf{L}'}(\mathbf{L}')|} + \alpha(\deg \mathfrak{B}', \|\mathfrak{B}'\|) = \frac{\|\mathfrak{B}'\|}{|\Phi_{\mathbf{L}'}(\overline{\mathbf{L}}^\ell)|^2} + \alpha(\deg \mathfrak{B}', \|\mathfrak{B}'\|) \leq \frac{2\|\mathfrak{B}'\|}{|\Phi_{\mathbf{L}'}(\overline{\mathbf{L}}^\ell)|^2}.$$

Applying equidistribution again we have that the first term is at most

$$\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)^2 \left(\frac{1}{\|\mathfrak{B}'\|} + \alpha(\deg \mathfrak{B}', \|\mathfrak{B}'\|) \right) \leq \frac{2\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)^2}{\|\mathfrak{B}'\|}.$$

Combining these bounds and summing over all terms that contain some f_{sml}^j , we see that

$$p_3 \geq \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in J} f^{(i)}(L_i^\ell(\mathbf{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \prod_{i \in E_\ell \setminus J} f_{\text{str}}^{(i)}(L_i^\ell(\mathbf{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \right] - 2^{|E_\ell|+1} \frac{\theta(\deg \mathfrak{B}, \|\mathfrak{B}\|)}{|\Phi_{\mathbf{L}'}(\overline{\mathbf{L}}^\ell)|}.$$

Write p_4 for the expectation above. The quantity $f_{\text{str}}^{(i)}(L_i^\ell(\mathbf{x}))$ is the density of $\psi(i)$ in the atom of \mathfrak{B}' that $L_i^\ell(\mathbf{x})$ lies in. When $\mathfrak{B}(L_i^\ell(\mathbf{x})) = s(a_i)$, the fact that $\psi(i)$ is high-density in a_i for all $i \in E_\ell \setminus J$ implies that

$$p_4 \geq \left(\frac{\epsilon}{4|\mathcal{S}|} \right)^{|E_\ell|} \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in J} f^{(i)}(L_i^\ell(\mathbf{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \prod_{i \in E_\ell \setminus J} 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \right].$$

Write p_5 for the expectation above. We write $\mathbf{L}_J := (L_i^\ell)_{i \in J}$. By assumption, we know that

$$(6.1) \quad \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in J} f^{(i)}(L_i^\ell(\mathbf{x})) 1_{\mathfrak{B}'^{-1}(s(a_i))}(L_i^\ell(\mathbf{x})) \right] \geq \frac{\beta(\deg I, \|I\|)}{|\Phi_{I'}(\mathbf{L}_J)|}.$$

We will use (6.1) to show that p_5 is at least on the order of $\beta(\deg I, \|I\|)/|\Phi_{I'}(\overline{\mathbf{L}}^\ell)|$. For simplicity, write $\beta := \beta(\deg I, \|I\|)$ in the rest of this argument.

By applying a change of coordinates, we can assume that \mathbf{L}_J only depends on $x_1, \dots, x_{\ell'}$ and is independent of $x_{\ell'+1}, \dots, x_\ell$. To lower bound p_5 , we want to show that each tuple $(x_1, \dots, x_{\ell'})$ that lies in certain atoms extends to a tuple (x_1, \dots, x_ℓ) that still lies in certain atoms in approximately the same number of ways. We do this by a Cauchy–Schwarz argument. Define \mathbf{L}'' to be the following system of $2|E_\ell| - |J|$ linear forms in $2\ell - \ell'$ variables. For $i \in J$, set

$$L_i''(x_1, \dots, x_\ell, x_{\ell'+1}, \dots, x_\ell') := L_i^{\ell'}(x_1, \dots, x_{\ell'}),$$

and for $i \in E_\ell \setminus J$, define

$$L_{i,1}''(x_1, \dots, x_\ell, x_{\ell'+1}, \dots, x_\ell') := L_i^\ell(x_1, \dots, x_\ell),$$

$$L_{i,2}''(x_1, \dots, x_\ell, x_{\ell'+1}, \dots, x_\ell') := L_i^\ell(x_1, \dots, x_{\ell'}, x_{\ell'+1}, \dots, x_\ell).$$

By Lemma 3.23, we know that $|\Phi_{I'}(\mathbf{L}_J)| \cdot |\Phi_{I'}(\mathbf{L}'')| = |\Phi_{I'}(\overline{\mathbf{L}}^\ell)|^2$.

Define $S \subseteq V^{\ell'}$ to be the set of tuples $\mathbf{x} = (x_1, \dots, x_{\ell'})$ such that $\mathfrak{B}'(L_i^\ell(\mathbf{x})) = s(a_i)$ for each $i \in J$. For $\mathbf{x} \in S$, let $c_{\mathbf{x}}$ be the number of tuples $\mathbf{x}' = (x_1, \dots, x_\ell)$ such that $\mathfrak{B}'(L_i^\ell(\mathbf{x})) = s(a_i)$ for each $i \in E_\ell$. By applying equidistribution (Theorem 3.19) to \mathbf{L}_J and $\overline{\mathbf{L}}^\ell$ and \mathbf{L}'' , we find that

$$(6.2) \quad |S| = \sum_{\mathbf{x} \in S} 1 \leq (1 + \beta/10) \frac{|V|^{\ell'}}{|\Phi_{I'}(\mathbf{L}_J)|},$$

$$(6.3) \quad \sum_{\mathbf{x} \in S} c_{\mathbf{x}} \geq (1 - \beta/10) \frac{|V|^\ell}{|\Phi_{I'}(\overline{\mathbf{L}}^\ell)|},$$

$$(6.4) \quad \sum_{\mathbf{x} \in S} c_{\mathbf{x}}^2 \leq (1 + \beta/10) \frac{|V|^{2\ell - \ell'}}{|\Phi_{I'}(\mathbf{L}'')|} = (1 + \beta/10) \frac{|V|^{2\ell - \ell'} |\Phi_{I'}(\mathbf{L}_J)|}{|\Phi_{I'}(\overline{\mathbf{L}}^\ell)|^2}.$$

Define $T \subseteq S \subseteq V^{\ell'}$ to be the set of tuples $\mathbf{x} = (x_1, \dots, x_{\ell'})$ such that $\mathfrak{B}'(L_i^\ell(\mathbf{x})) = s(a_i)$ and $f^{(i)}(L_i^\ell(\mathbf{x})) = 1$ for each $i \in J$. (6.1) implies that

$$(6.5) \quad |T| \geq \beta \frac{|V|^{\ell'}}{|\Phi_{I'}(\mathbf{L}_J)|}.$$

We express p_5 as follows:

$$p_5 = \frac{1}{|V|^\ell} \sum_{\mathbf{x} \in T} c_{\mathbf{x}} = \frac{1}{|V|^\ell} \left(\sum_{\mathbf{x} \in S} c_{\mathbf{x}} - \sum_{\mathbf{x} \in S \setminus T} c_{\mathbf{x}} \right) \geq (1 - \beta/10) \frac{1}{|\Phi_{I'}(\mathbf{L}_J)|} - \frac{1}{|V|^\ell} \sum_{\mathbf{x} \in S \setminus T} c_{\mathbf{x}}.$$

Then combining (6.2), (6.4), (6.5) with the Cauchy–Schwarz inequality gives

$$\left(\sum_{\mathbf{x} \in S \setminus T} c_{\mathbf{x}} \right)^2 \leq |S \setminus T| \cdot \sum_{\mathbf{x} \in S \setminus T} c_{\mathbf{x}}^2 \leq (1 - 4\beta/5) \frac{|V|^{2\ell}}{|\Phi_{I'}(\overline{\mathbf{L}}^\ell)|^2}.$$

Taking the square root and combining the above two inequalities gives

$$p_5 \geq \frac{\beta}{10} \frac{1}{|\Phi_{I'}(\overline{\mathbf{L}}^\ell)|}.$$

Combining all the above inequalities we see that p_1 , the H -density in f , is bounded by

$$p_1 \geq \left(\left(\frac{\epsilon}{4|S|} \right)^{|E_\ell|} \frac{\beta(\deg \mathfrak{B}, \|\mathfrak{B}\|)}{10} - 2^{|E_\ell|+1} \theta(\deg \mathfrak{B}, \|\mathfrak{B}\|) \right) \frac{1}{|\Phi_{I'}(\overline{\mathbf{L}}^\ell)|} - 3^{|E_\ell|} \eta(\deg \mathfrak{B}', \|\mathfrak{B}'\|) > \delta(\epsilon, \mathcal{H}).$$

This provides the desired contradiction. Therefore we conclude that the recoloring $g: V \rightarrow \mathcal{S}$ has no generic H -instances for every $H \in \mathcal{H}$. \square

We are finally in a position to prove the main removal lemma. To do so we take an arbitrary coloring $f: V \rightarrow \mathcal{S}$ and construct a new projective coloring \bar{f} where the color of x under \bar{f} contains the information of the coloring of the whole line spanned by x . Once we have done so, we can deduce the main removal lemma from the projective removal lemma.

Proof of Theorem 2.4. Define $\overline{\mathcal{S}} := \mathcal{S}^{\mathbb{F}_p^\times}$ with \mathbb{F}_p^\times -action defined by

$$b' \cdot (c_b)_{b \in \mathbb{F}_p^\times} := (c_{b'b})_{b \in \mathbb{F}_p^\times}.$$

First we partition

$$\mathcal{H} = \bigsqcup_{c \in \mathcal{S} \sqcup \{0\}} \mathcal{H}_c$$

as follows. If $L_i \equiv 0$ for some $i \in [m]$, place H in the set $\mathcal{H}_{\psi(i)}$. Otherwise, place H in \mathcal{H}_0 . (Without loss of generality, we can assume that no pattern in \mathcal{H} has multiple linear forms that are identically equal to 0.)

Now we define sets $\overline{\mathcal{H}}_c$ of $\overline{\mathcal{S}}$ -colored patterns for each $c \in \mathcal{S}$. Let $H = (\mathbf{L}, \psi) \in \mathcal{H}_0 \cup \mathcal{H}_c$ be an \mathcal{S} -colored pattern over \mathbb{F}_p consisting of m linear forms in ℓ variables. We can write $\mathbf{L} = (L_i^\ell)_{i \in J}$ for some set $J \subseteq \mathbb{F}_p^\ell$ of size m . We convert H to a set of patterns defined by the system $\overline{\mathbf{L}}^\ell = (L_i^\ell)_{i \in E_\ell}$ as follows. (See Definition 2.7 for the definitions of L_i^ℓ and E_ℓ .) For each function $\bar{\psi}: E_\ell \rightarrow \overline{\mathcal{S}}$ that satisfies $\bar{\psi}(i)_b = \psi(bi)$ whenever $i \in E_\ell \subseteq \mathbb{F}_p^\ell$ and $b \in \mathbb{F}_p^\times$ are such that $bi \in J \subseteq \mathbb{F}_p^\ell$, we add $(\overline{\mathbf{L}}^\ell, \bar{\psi})$ to $\overline{\mathcal{H}}_c$.

For each $c \in \mathcal{S}$, we apply Theorem 2.8 to $\overline{\mathcal{H}}_c$ with parameter ϵ . This produces a finite subset $\overline{\mathcal{H}}_{c,\epsilon} \subseteq \overline{\mathcal{H}}_c$ and $\delta_c = \delta(\epsilon, \overline{\mathcal{H}}_c) > 0$ with several desirable properties.

Let $\mathcal{H}_\epsilon \subseteq \mathcal{H}$ be the finite subset consisting of all patterns H such that some pattern \bar{H} corresponding to H lies in $\overline{\mathcal{H}}_{c,\epsilon}$ for some $c \in \mathcal{S}$. Let $\delta = \min_{c \in \mathcal{S}} \delta_c > 0$. We claim that $\mathcal{H}_\epsilon, \delta$ satisfy the desired conclusion.

Let V be a finite-dimensional \mathbb{F}_p -vector space. Let $f: V \rightarrow \mathcal{S}$ be a function. Suppose that the H -density in f is at most δ for every $H \in \mathcal{H}_\epsilon$. Define $\bar{f}: V \rightarrow \overline{\mathcal{S}}$ by

$$\bar{f}(x) := (f(bx))_{b \in \mathbb{F}_p^\times}.$$

Note that \bar{f} is a projective function. Furthermore, we claim that the \bar{H} -density in \bar{f} is at most δ for all $\bar{H} \in \bar{\mathcal{H}}_{f(0)}$. This is true simply because if $\mathbf{x} = (x_1, \dots, x_\ell) \in V^\ell$ is an \bar{H} -instance in \bar{f} , then \mathbf{x} is also an H -instance in f where $\bar{H} \in \bar{\mathcal{H}}_{f(0)}$ is any pattern corresponding to $H \in \mathcal{H}_0 \sqcup \mathcal{H}_{f(0)}$.

Thus by assumption, there exists a projective recoloring $\bar{g}: V \rightarrow \bar{\mathcal{S}}$ such that \bar{g} agrees with \bar{f} on all but an at most ϵ -fraction of V and \bar{g} has no generic \bar{H} -instances for every $\bar{H} \in \bar{\mathcal{H}}_{f(0)}$.

Define $g: V \rightarrow \mathcal{S}$ by $g(x) := \bar{g}(x)_1$ for $x \neq 0$ and $g(0) = f(0)$. Note that g agrees with f on all but an at most ϵ -fraction of V . Furthermore, note that g has no H -instances for $H \in \mathcal{H}_c$ with $c \neq f(0)$ since $f(0) = g(0)$. Finally, g has no generic H -instances for $H \in \mathcal{H}_0 \sqcup \mathcal{H}_{f(0)}$ since any such generic H -instance in g is a generic \bar{H} -instance in \bar{g} for some \bar{H} corresponding H in $\bar{\mathcal{H}}_{f(0)}$, which we assumed was not the case. \square

7. Proof of property testing results. In this section we prove our two property testing results, Theorems 1.3 and 1.4 from the main removal lemma.

Proof of Theorem 1.3. It follows from [9, Theorem 10] that a linear-invariant property is testable only if it is semi-subspace-hereditary.

Now suppose that \mathcal{P} is a linear-invariant semi-subspace-hereditary property. By definition, there exists a subspace-hereditary property \mathcal{Q} such that

- (i) every function satisfying \mathcal{P} also satisfies \mathcal{Q} ;
- (ii) for all $\epsilon > 0$, there exists $N(\epsilon)$ such that if $f: V \rightarrow \mathcal{S}$ satisfies \mathcal{Q} and is ϵ -far from satisfying \mathcal{P} , then $\dim V < N(\epsilon)$.

We define \mathcal{H} a (possibly infinite) set of \mathcal{S} -colored patterns. For each $f: \mathbb{F}_p^\ell \rightarrow \mathcal{S}$ that does not satisfy \mathcal{Q} , include $H = (\mathbf{L}^\ell, f)$ in \mathcal{H} (\mathbf{L}^ℓ is the system of linear forms that defines an ℓ -dimensional subspace, defined in Definition 2.7). Since \mathcal{Q} is subspace-hereditary, it immediately follows that \mathcal{Q} consists exactly of the functions with no generic H -instances for any $H \in \mathcal{H}$.

By Theorem 2.4, there exist a finite subset $\mathcal{H}_\epsilon \subseteq \mathcal{H}$ and some $\delta(\epsilon, \mathcal{H}) > 0$ such that the following holds. If $f: V \rightarrow \mathcal{S}$ has H -density at most $\delta(\epsilon, \mathcal{H})$ for every $H \in \mathcal{H}_\epsilon$, then f is ϵ -close to \mathcal{Q} . Define $\ell(\epsilon)$ to be the largest ℓ such that some pattern defined by the system \mathbf{L}^ℓ is present in \mathcal{H}_ϵ .

Now we define the oblivious tester for \mathcal{P} . Given $\epsilon > 0$, the tester produces

$$d(\epsilon) := \max \{N(\epsilon/2), \lceil \log_p(2/\delta(\epsilon/2, \mathcal{H})) \rceil + \ell(\epsilon/2)\}.$$

Given a function $f: V \rightarrow \mathcal{S}$ our tester receives oracle access to $f|_U$ where

- (i) if $\dim V \geq d(\epsilon)$, then U is a uniform random affine subspace of dimension $d(\epsilon)$;
- (ii) else, $U = V$.

Our tester works as follows. If $\dim U < d(\epsilon)$ the tester accepts if $f|_U \in \mathcal{P}$. If $\dim U \geq d(\epsilon)$ the tester accepts if $f|_U \in \mathcal{Q}$.

Suppose $f \in \mathcal{P}$. If $\dim U < d(\epsilon)$, then $U = V$, so $f|_U = f \in \mathcal{P}$. Thus the tester always accepts in this case. In the other case, note that since $f \in \mathcal{P}$, it follows that $f \in \mathcal{Q}$, and since \mathcal{Q} is subspace-hereditary, $f|_U \in \mathcal{Q}$. Thus the tester also always accepts in this case.

Now suppose that f is ϵ -far from \mathcal{P} . By the definition of \mathcal{Q} we know that either $\dim V < N(\epsilon/2) \leq d(\epsilon)$ or f is $\epsilon/2$ -far from \mathcal{Q} . Consider the action of the tester. If

$\dim U < d(\epsilon)$, then $U = V$ so $f|_U = f \notin \mathcal{P}$. Thus the tester always rejects in this case. In the other case, note that since f is $\epsilon/2$ -far from \mathcal{Q} , by assumption there is some $H \in \mathcal{H}_{\epsilon/2}$ such that f has H -density more than $\delta(\epsilon/2, \mathcal{H})$. Let $H = (\mathbf{L}^\ell, \psi)$ for some $\ell \leq \ell(\epsilon/2)$. We claim that the fact that the H -density in f is large implies that there is at least a $\delta(\epsilon/2, \mathcal{H})/2$ -fraction of ℓ -dimensional subspaces that f colors by ψ . (Note that this does not immediately follow since the H -density includes the contribution of H -instances that are not generic.) We can compute that the probability a uniform random \mathbf{L}^ℓ -instance in V is not generic is at most $p^{\ell - \dim V}$. It follows that the fraction of ℓ -dimensional subspaces that f colors by ψ is at least

$$\delta(\epsilon/2, \mathcal{H}) - p^{\ell - \dim V} \geq \delta(\epsilon/2, \mathcal{H}) - p^{-\lceil \log_p(2/\delta(\epsilon/2, \mathcal{H})) \rceil} \geq \delta(\epsilon/2, \mathcal{H})/2.$$

We conclude that in this case the tester rejects with probability at least $\delta(\epsilon/2, \mathcal{H})/2$, as desired. \square

Proof of Theorem 1.4. Suppose \mathcal{P} is a linear-invariant property that is PO-testable. By definition, there exists some d , independent of ϵ and $\dim V$ such that to test $f: V \rightarrow \mathcal{S}$, the tester receives $f|_U$ where U is

- (i) if $\dim V \geq d$, then U is a uniform random linear subspace of dimension d ;
- (ii) else, $U = V$.

We define \mathcal{H} to be the set of patterns of the form (\mathbf{L}^d, ψ) where $\psi: \mathbb{F}_p^d \rightarrow \mathcal{S}$ is a restriction that the tester rejects on $(\mathbf{L}^d$ is the pattern that defines a d -dimensional subspace, defined in Definition 2.7). We claim that for every $f: V \rightarrow \mathcal{S}$ with $\dim V \geq d$, it holds that $f \in \mathcal{P}$ if and only if f has no generic \mathcal{H} -instances. This claim suffices to prove that \mathcal{H} is subspace-hereditary and locally characterized.

Suppose $f: V \rightarrow \mathcal{S}$ satisfies \mathcal{P} and $\dim V \geq d$. By the definition of PO-testable, the tester must accept f with probability 1. Thus the tester must accept $f|_U$ for every $U \leq V$ of dimension d . This implies that f has no generic \mathcal{H} -instances. Now suppose that $f: V \rightarrow \mathcal{S}$ does not satisfy \mathcal{P} and $\dim V \geq d$ holds. By definition, the tester must accept f with positive probability. Thus there must be some $U \leq V$ of dimension d such that $f|_U$ rejects. This is equivalent to the fact that f contains a generic H -instance for some $H \in \mathcal{H}$, proving the desired result.

Now we show that every linear-invariant subspace-hereditary locally characterized property is testable. Suppose \mathcal{P} is such a property. It follows that there is some d and (finite) \mathcal{H} consisting of patterns of the form (\mathbf{L}^d, ψ) such that for $f: V \rightarrow \mathcal{S}$ with $\dim V \geq d$, we have f satisfies \mathcal{P} if and only if f has no generic \mathcal{H} -instances.

The PO-tester for \mathcal{P} proceeds in the obvious way. The tester is given oracle access to $f|_U$ where U is

- (i) if $\dim V \geq d$, then U is a uniform random linear subspace of dimension d ;
- (ii) else, $U = V$.

The tester accepts if and only if $f|_U \in \mathcal{P}$.

Suppose $f: V \rightarrow \mathcal{S}$ satisfies \mathcal{P} . If $\dim V < d$, then $f|_U = f \in \mathcal{P}$, so the tester accepts f . If $\dim V \geq d$, then by the fact that \mathcal{P} is subspace-hereditary and locally characterized, it follows that $f|_U \in \mathcal{P}$ for all d -dimensional $U \leq V$. Thus the tester accepts f in this case as well.

Now suppose that $f: V \rightarrow \mathcal{S}$ is ϵ -far from \mathcal{P} . If $\dim V < d$, then $f|_U = f \notin \mathcal{P}$, so the tester rejects f . If $\dim V \geq d$, by Theorem 2.4, there must be some $H = (\mathbf{L}^d, \psi) \in \mathcal{H}$ such that the H -density in f is more than $\delta(\epsilon, \mathcal{H})$. We claim that this implies that there is a large fraction of d -dimensional subspaces that f colors by some

$H \in \mathcal{H}$. (Note that this does not immediately follow since the H -density includes the contribution of H -instances that are not generic.) We can compute that at most a $p^{d-\dim V}$ -fraction of H -instances are nongeneric. Thus at least the fraction of d -dimensional subspaces that f colors by ψ is at least

$$\delta(\epsilon, \mathcal{H}) - p^{d-\dim V}.$$

This is a good bound on the rejection probability when $\dim V$ is large, but it can be quite bad (or even negative) when $\dim V$ is small. In the latter case we use the following easy bound. If f does not satisfy \mathcal{P} , there is at least one d -dimensional subspace that is colored by \mathcal{H} . The probability that the tester finds this one \mathcal{H} -instance is lower-bounded by $p^{-d-\dim V}$. Thus the rejection probability of this tester is at least

$$\max \{ \delta(\epsilon) - p^{d-\dim V}, p^{-d-\dim V} \}.$$

Note that for $\dim V \geq d$, this quantity is uniformly bounded away from 0 (independent of $\dim V$), as desired. \square

Acknowledgments. We thank the anonymous referees for many comments which improved the exposition of this paper.

REFERENCES

- [1] N. ALON, R. A. DUKE, H. LEFMAN, V. RÖDL, AND R. YUSTER, *The algorithmic aspects of the regularity lemma*, J. Algorithms, 16 (1994), pp. 80–109.
- [2] N. ALON, E. FISCHER, M. KRIVELEVICH, AND M. SZEGEDY, *Efficient testing of large graphs*, Combinatorica, 20 (2000), pp. 451–476.
- [3] N. ALON, T. KAUFMAN, M. KRIVELEVICH, S. LITSYN, AND D. RON, *Testing Reed-Muller codes*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4032–4039.
- [4] N. ALON AND A. SHAPIRA, *A characterization of the (natural) graph properties testable with one-sided error*, SIAM J. Comput., 37 (2008), pp. 1703–1727.
- [5] V. BERGELSON, T. TAO, AND T. ZIEGLER, *An inverse theorem for the uniformity seminorms associated with the action of \mathbb{F}_p^∞* , Geom. Funct. Anal., 19 (2010), pp. 1539–1596.
- [6] A. BHATTACHARYYA, V. CHEN, M. SUDAN, AND N. XIE, *Testing linear-invariant non-linear properties*, Theory Comput., 7 (2011), pp. 75–99.
- [7] A. BHATTACHARYYA, E. FISCHER, H. HATAMI, P. HATAMI, AND S. LOVETT, *Every locally characterized affine-invariant property is testable*, in STOC’13—Proceedings of the 2013 ACM Symposium on Theory of Computing, ACM, New York, 2013, pp. 429–435.
- [8] A. BHATTACHARYYA, E. FISCHER, AND S. LOVETT, *Testing low complexity affine-invariant properties*, in Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, 2012, pp. 1337–1355.
- [9] A. BHATTACHARYYA, E. GRIGORESCU, AND A. SHAPIRA, *A unified framework for testing linear-invariant properties*, Random Structures Algorithms, 46 (2015), pp. 232–260.
- [10] M. BLUM, M. LUBY, AND R. RUBINFELD, *Self-testing/correcting with applications to numerical problems*, J. Comput. System Sci., 47 (1993), pp. 549–595.
- [11] D. CONLON AND J. FOX, *Graph removal lemmas*, in surveys in Combinatorics 2013, London Math. Soc. Lecture Note Ser. 409, Cambridge University Press, Cambridge, UK, 2013, pp. 1–49.
- [12] J. FOX, J. TIDOR, AND Y. ZHAO, *Induced arithmetic removal: Complexity 1 patterns over finite fields*, Israel J. Math., 248 (2022), pp. 1–38.
- [13] Z. FÜREDI, *Extremal hypergraphs and combinatorial geometry*, in Proceedings of the International Congress of Mathematicians, Birkhäuser, Basel, 1995, pp. 1343–1352.
- [14] O. GOLDBREICH, S. GOLDWASSER, AND D. RON, *Property testing and its connection to learning and approximation*, J. ACM, 45 (1998), pp. 653–750.
- [15] O. GOLDBREICH AND D. RON, *On proximity-oblivious testing*, SIAM J. Comput., 40 (2011), pp. 534–566.
- [16] W. TIMOTHY GOWERS, *A new proof of Szemerédi’s theorem*, Geom. Funct. Anal., 11 (2001), pp. 465–588.

- [17] W. T. GOWERS, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2), 166 (2007), pp. 897–946.
- [18] W. T. GOWERS AND J. WOLF, *Linear forms and higher-degree uniformity for functions on \mathbb{F}_p^n* , Geom. Funct. Anal., 21 (2011), pp. 36–69.
- [19] B. GREEN, *A Szemerédi-type regularity lemma in abelian groups, with applications*, Geom. Funct. Anal., 15 (2005), pp. 340–376.
- [20] B. GREEN AND T. SANDERS, *Fourier Uniformity on Subspaces*, arXiv:1510.08739, 2015.
- [21] B. GREEN AND T. TAO, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2), 167 (2008), pp. 481–547.
- [22] B. GREEN AND T. TAO, *Linear equations in primes*, Ann. of Math. (2), 171 (2010), pp. 1753–1850.
- [23] B. GREEN, T. TAO, AND T. ZIEGLER, *An inverse theorem for the Gowers $U^{s+1}[N]$ -norm*, Ann. of Math. (2), 176 (2012), pp. 1231–1372.
- [24] H. HATAMI, P. HATAMI, AND S. LOVETT, *General systems of linear forms: Equidistribution and true complexity*, Adv. Math., 292 (2016), pp. 446–477.
- [25] H. HATAMI, P. HATAMI, AND S. LOVETT, *Higher-order Fourier analysis and applications*, Found. Trends Theor. Comput. Sci., 13 (2019), pp. 247–448.
- [26] T. KAUFMAN AND M. SUDAN, *Algebraic property testing: The role of invariance*, in Proceedings of STOC’08, ACM, New York, 2008, pp. 403–412.
- [27] D. KRÁL’, O. SERRA, AND L. VENA, *A removal lemma for systems of linear equations over finite fields*, Israel J. Math., 187 (2012), pp. 193–207.
- [28] V. RÖDL, B. NAGLE, J. SKOKAN, M. SCHACHT, AND Y. KOHAYAKAWA, *The hypergraph regularity method and its applications*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 8109–8113.
- [29] V. RÖDL AND M. SCHACHT, *Property testing in hypergraphs and the removal lemma [extended abstract]*, in STOC’07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 488–495.
- [30] R. RUBINFELD AND M. SUDAN, *Robust characterizations of polynomials with applications to program testing*, SIAM J. Comput., 25 (1996), pp. 252–271.
- [31] I. Z. RUZSA AND E. SZEMERÉDI, *Triple systems with no six points carrying three triangles*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai 18, North-Holland, Amsterdam, 1978, pp. 939–945.
- [32] A. SHAPIRA, *A proof of Green’s conjecture regarding the removal properties of sets of linear equations*, J. Lond. Math. Soc. (2), 81 (2010), pp. 355–373.
- [33] T. TAO AND T. ZIEGLER, *The inverse conjecture for the Gowers norm over finite fields via the correspondence principle*, Anal. PDE, 3 (2010), pp. 1–20.
- [34] T. TAO AND T. ZIEGLER, *The inverse conjecture for the Gowers norm over finite fields in low characteristic*, Ann. Comb., 16 (2012), pp. 121–188.