# A polynomial regularity lemma for semi-algebraic hypergraphs and its applications in geometry and property testing[*]

Jacob Fox [†]     János Pach[‡]     Andrew Suk[§]

February 9, 2015

### Abstract

Fox, Gromov, Lafforgue, Naor, and Pach proved a regularity lemma for semi-algebraic $k$-uniform hypergraphs of bounded complexity, showing that for each $\epsilon > 0$ the vertex set can be equitably partitioned into a bounded number of parts (in terms of $\epsilon$ and the complexity) so that all but an $\epsilon$-fraction of the $k$-tuples of parts are homogeneous. We prove that the number of parts can be taken to be polynomial in $1/\epsilon$. Our improved regularity lemma can be applied to geometric problems and to the following general question on property testing: is it possible to decide, with query complexity polynomial in the reciprocal of the approximation parameter, whether a hypergraph has a given hereditary property? We give an affirmative answer for testing typical hereditary properties for semi-algebraic hypergraphs of bounded complexity.

## 1 Introduction

A $k$-uniform hypergraph $H = (P, E)$ consists of a vertex set $P$ and an edge set (or hyperedge set) $E$, which is a collection of $k$-element subsets of $P$. A hypergraph is *$k$-partite* if it is $k$-uniform and its vertex set $P$ is partitioned into $k$ parts, $P = P_1 \cup \ldots \cup P_k$, such that every edge has precisely one vertex in each part. It follows from a classical theorem of Erdős [17], which was one of the starting points in extremal hypergraph theory, that if $|P_1| = \ldots = |P_k|$ and $|E| \geq \varepsilon \Pi_{i=1}^k |P_i|$ for some $\varepsilon > 0$, then one can find subsets $P_i' \subset P_i$ such that

$$|P_i'| = \Omega \left( \frac{\log |P_i|}{\log(1/\varepsilon)} \right)^{1/(k-1)},$$

and $P_1' \times \cdots \times P_k' \subset E$. In other words, $H$ contains a large *complete $k$-partite subhypergraph*. For graphs ($k = 2$), this was already shown in [32].

It turns out that much larger complete $k$-partite subhypergraphs can be found in hypergraphs that admit a simple algebraic description. To make this statement precise, we need some terminology.

---

**Semi-algebraic setting.** A $k$-partite hypergraph $H = (P_1 \cup \ldots \cup P_k, E)$ is called *semi-algebraic* in $\mathbb{R}^d$, if its vertices are points in $\mathbb{R}^d$, and there are polynomials $f_1, \ldots, f_t \in \mathbb{R}[x_1, \ldots, x_{kd}]$ and a Boolean function $\Phi$ such that for every $p_i \in P_i$ $(i = 1, \ldots, k)$, we have

$$(p_1, \ldots, p_k) \in E \quad \Leftrightarrow \quad \Phi(f_1(p_1, \ldots, p_k) \geq 0; \ldots; f_t(p_1, \ldots, p_k) \geq 0) = 1.$$

At the evaluation of $f_j(p_1, \ldots, p_k)$, we substitute the variables $x_1, \ldots, x_k$ with the coordinates of $p_1$, the variables $x_{k+1}, \ldots, x_{2k}$ with the coordinates of $p_2$, etc.

We say that $H$ has *complexity* $(t, D)$ if each polynomial $f_j$ with $1 \leq j \leq t$ has the property that for any fixed $k - 1$ points $q_1, \ldots, q_{k-1} \in \mathbb{R}^d$, the $d$-variate polynomials

$$
\begin{aligned}
h_{j,1}(\mathbf{x}_1) &= f_j(\mathbf{x}_1, q_1, \ldots, q_{k-1}), \\
h_{j,2}(\mathbf{x}_2) &= f_j(q_1, \mathbf{x}_2, q_2, \ldots, q_{k-1}), \\
&\vdots \\
h_{j,k}(\mathbf{x}_k) &= f_j(q_1, \ldots, q_{k-1}, \mathbf{x}_k),
\end{aligned}
$$

are of degree at most $D$ (in notation, $deg(h_{j,i}) \leq D$ for $1 \leq j \leq t$ and $1 \leq i \leq k$). It follows that $deg(f_j) \leq kD$ for every $j$.

If our $k$-uniform hypergraph $H = (P, E)$ is *a priori* not $k$-partite, we fix an enumeration $p_1, p_2, \ldots$ of the elements of $P \subset \mathbb{R}^d$, and we say that $H$ is *semi-algebraic* with complexity $(t, D)$ if for every $1 \leq i_1 < \cdots < i_k \leq n$,

$$(p_{i_1}, \ldots, p_{i_k}) \in E \quad \Leftrightarrow \quad \Phi(f_1(p_{i_1}, \ldots, p_{i_k}) \geq 0, \ldots, f_t(p_{i_1}, \ldots, p_{i_k}) \geq 0) = 1,$$

where $\Phi$ is a Boolean function and $f_1, \ldots, f_t$ are polynomials satisfying the same properties as above.

**Density theorem for semi-algebraic hypergraphs.** Fox *et al.* [19] showed that there exists a constant $c = c(k, d, t, D) > 0$ with the following property. Let $(P_1 \cup \ldots \cup P_k, E)$ be any $k$-partite semi-algebraic hypergraph in $\mathbb{R}^d$ with complexity $(t, D)$, and suppose that $|E| \geq \varepsilon \Pi_{i=1}^k |P_i|$. Then one can find subsets $P_i' \subset P_i$, $1 \leq i \leq k$, with $|P_i'| \geq \varepsilon^c |P_i|$ which induce a complete $k$-partite subhypergraph, that is, $P_1' \times \ldots \times P_k' \subset E$. The original proof gives a poor upper bound on $c(k, d, t, D)$, which is tower-type in $k$. Combining a result of Bukh and Hubard [11] with a variational argument of Komlós [31], [34], the dependence on $k$ can be improved to double exponential. Our following result, which will be proved in Section 2, reduces this bound to linear in $k$.

**Theorem 1.1.** *For any positive integers $d, t, D$, there exists a constant $C = C(d, t, D)$ with the following property. Let $\varepsilon > 0$ and let $H = (P_1 \cup \ldots \cup P_k, E)$ be any $k$-partite semi-algebraic hypergraph in $\mathbb{R}^d$ with complexity $(t, D)$ and $|E| \geq \varepsilon \Pi_{i=1}^k |P_i|$. Then one can choose subsets $P_i' \subset P_i$, $1 \leq i \leq k$, such that*

$$|P_i'| \geq \frac{\varepsilon^{\binom{d+D}{d}}}{C^k} |P_i|,$$

*and $P_1' \times \ldots \times P_k' \subseteq E$. Moreover, we can take $C = 2^{20m \log(m+1)} t^{m/k}$, where $m = \binom{d+D}{d} - 1$.*

In many applications, the dependency on the dimension $d$ also becomes crucial. This is typically the case for relations that have complexity $(t, 1)$ (i.e., when $D = 1$).

Substituting $D = 1$ in Theorem 1.1, we obtain the following.

**Corollary 1.2.** *Let $\varepsilon > 0$ and let $H = (P_1 \cup \ldots \cup P_k, E)$ be a $k$-partite semi-algebraic hypergraph in $\mathbb{R}^d$ with complexity $(t, 1)$ and $|E| \geq \varepsilon \Pi_{i=1}^{k} |P_i|$. Then one can choose subsets $P_i' \subset P_i$, $1 \leq i \leq k$, such that*

$$|P_i'| \geq \frac{\varepsilon^{d+1}}{2^{20kd\log(d+1)}t^d}|P_i|,$$

*and $P_1' \times \cdots \times P_k' \subseteq E$.*

**Polynomial regularity lemma for semi-algebraic hypergraphs.** Szemerédi's regularity lemma is one of the most powerful tools in modern combinatorics. In its simplest version [42] it gives a rough structural characterization of all graphs. A partition is called *equitable* if any two parts differ in size by at most one. According to the lemma, for every $\varepsilon > 0$ there is $K = K(\varepsilon)$ such that every graph has an equitable vertex partition into at most $K$ parts such that all but at most an $\varepsilon$ fraction of the pairs of parts behave "regularly".[1] The dependence of $K$ on $1/\varepsilon$ is notoriously bad. It follows from the proof that $K(\varepsilon)$ may be taken to be of an exponential tower of 2-s of height $\varepsilon^{-O(1)}$. Gowers [26] used a probabilistic construction to show that such an enormous bound is indeed necessary. Consult [15], [35], [21] for other proofs that improve on various aspects of the result. Szemerédi's regularity lemma was extended to $k$-uniform hypergraphs by Gowers [25, 27] and by Nagle, Rödl, Schacht, and Skokan [36]. The bounds on the number of parts go up in the Ackermann hierarchy, as $k$ increases. This is quite unfortunate, because in property testing and in other algorithmic applications of the regularity lemma this parameter has a negative impact on the efficiency.

Alon *et al.* [3] (for $k = 2$) and Fox *et al.* [19] (for $k > 2$) established an "almost perfect" regularity lemma for $k$-uniform semi-algebraic hypergraphs $H = (P, E)$. According to this, $P$ has an equitable partition such that all but at most an $\varepsilon$-fraction of the $k$-tuples of parts $(P_{i_1}, \ldots, P_{i_k})$ behave not only regularly, but *homogeneously* in the sense that either $P_{i_1} \times \ldots \times P_{i_k} \subseteq E$ or $P_{i_1} \times \ldots \times P_{i_k} \cap E = \emptyset$. The proof is essentially qualitative: it gives a very poor estimate for the number of parts in such a partition.

In Section 4, we deduce a much better quantitative form of this result, showing that the number of parts can be taken to be polynomial in $1/\varepsilon$.

**Theorem 1.3.** *For any positive integers $k, d, t, D$ there exists a constant $c = c(k, d, t, D) > 0$ with the following property. Let $0 < \varepsilon < 1/2$ and let $H = (P, E)$ be a $k$-uniform semi-algebraic hypergraph in $\mathbb{R}^d$ with complexity $(t, D)$. Then $P$ has an equitable partition $P = P_1 \cup \cdots \cup P_K$ into at most $K \leq (1/\varepsilon)^c$ parts such that all but an $\varepsilon$-fraction of the $k$-tuples of parts are homogeneous.*

See [16], for other favorable Ramsey-type properties of semi-algebraic sets.

**Geometric applications.** In Section 3, we prove three geometric applications of Corollary 1.2.

*1. Same-type lemma.* Let $P = (p_1, \ldots, p_n)$ be an $n$-element point sequence in $\mathbb{R}^d$ in *general position*, i.e., assume that no $d + 1$ points lie in a common hyperplane. For $i_1 < i_2 < \cdots < i_{d+1}$, the *orientation* of the $(d+1)$-tuple $(p_{i_1}, p_{i_2}, \ldots, p_{i_{d+1}}) \subset P$ is defined as the sign of the determinant of the unique linear mapping $M$ that sends the $d$ vectors $p_{i_2} - p_{i_1}, p_{i_3} - p_{i_1}, \ldots, p_{i_{d+1}} - p_{i_1}$, to

---

[1]For a pair $(P_i, P_j)$ of vertex subsets, $e(P_i, P_j)$ denotes the number of edges in the graph running between $P_i$ and $P_j$. The density $d(P_i, P_j)$ is defined as $\frac{e(P_i, P_j)}{|P_i||P_j|}$. The pair $(P_i, P_j)$ is called $\varepsilon$-regular if for all $P_i' \subset P_i$ and $P_j' \subset P_j$ with $|P_i'| \geq \varepsilon|P_i|$ and $|P_j'| \geq \varepsilon|P_j|$, we have $|d(P_i', P_j') - d(P_i, P_j)| \leq \varepsilon$.

the standard basis $e_1, e_2, \ldots, e_d$. Denoting the coordinates of $p_i$ by $x_{i,1}, \ldots, x_{i,d}$, the orientation of $(p_{i_1}, p_{i_2}, \ldots, p_{i_{d+1}})$ is

$$\chi = \operatorname{sgn} \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{d+1,1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,d} & x_{2,1} & \cdots & x_{d+1,1} \end{pmatrix}.$$

The *order-type* of $P = (p_1, p_2, \ldots, p_n)$ is the mapping $\chi : \binom{P}{d+1} \to \{+1, -1\}$ (positive orientation, negative orientation), assigning each $(d+1)$-tuple of $P$ its orientation. Therefore, two $n$-element point sequences $P$ and $Q$ have the same order-type if they are "combinatorially equivalent." See [34] and [23] for more background on order-types.

Let $(P_1, \ldots, P_k)$ be a $k$-tuple of finite sets in $\mathbb{R}^d$. A *transversal* of $(P_1, \ldots, P_k)$ is a $k$-tuple $(p_1, \ldots, p_k)$ such that $p_i \in P_i$ for all $i$. We say that $(P_1, \ldots, P_k)$ has *same-type transversals* if all of its transversals have the same order-type. Bárány and Valtr [9] showed that for $d, k > 1$, there exists a $c = c(d, k)$ such that the following holds. Let $P_1, \ldots, P_k$ be finite sets in $\mathbb{R}^d$ such that $P_1 \cup \cdots \cup P_k$ is in general position. Then there are subsets $P_1' \subset P_1, \ldots, P_k' \subset P_k$ such that the $k$-tuple $(P_1', \ldots, P_k')$ has same-type transversals and $|P_i'| \geq c(d, k)|P_i|$. Their proof shows that $c(d, k) = 2^{-k^{O(d)}}$. We make the following improvement.

**Theorem 1.4.** *For $k > d$, let $P_1, \ldots, P_k$ be finite sets in $\mathbb{R}^d$ such that $P_1 \cup \cdots \cup P_k$ is in general position. Then there are subsets $P_1' \subset P_1, \ldots, P_k' \subset P_k$ such that the $k$-tuple $(P_1', \ldots, P_k')$ has same-type transversals and*

$$|P_i'| \geq 2^{-O(d^3 k \log k)}|P_i|,$$

*for all $i$.*

*2. Homogeneous selections from hyperplanes.* Bárány and Pach [8] proved that for every integer $d \geq 2$, there is a constant $c = c(d) > 0$ with the following property. Given finite families $L_1, \ldots, L_{d+1}$ of hyperplanes in $\mathbb{R}^d$ in general position,[2] there are subfamilies $L_i' \subset L_i$ with $|L_i'| \geq c(d)|L_i|$, $1 \leq i \leq d + 1$, and a point $q \in \mathbb{R}^d$ such that for $(h_1, \ldots, h_{d+1}) \in L_1' \times \cdots \times L_{d+1}'$, $q$ lies in the unique bounded simplex $\Delta(h_1, \ldots, h_{d+1})$ enclosed by $\bigcup_{i=1}^{d+1} h_i$. The proof gives that one can take $c(d) = 2^{-(d+1)2^d}$, because they showed that $c(d) = c(d, d+2)$ will meet the requirements, where $c(d, k)$ denotes the constant defined above for same-type transversals. Thus, Theorem 1.4 immediately implies the following improvement.

**Theorem 1.5.** *Given finite families $L_1, \ldots, L_{d+1}$ of hyperplanes in $\mathbb{R}^d$ in general position, there are subfamilies $L_i' \subset L_i$, $1 \leq i \leq d + 1$, with*

$$|L_i'| \geq 2^{-O(d^4 \log(d+1))}|L_i|,$$

*and a point $q \in \mathbb{R}^d$ such that for every $(h_1, \ldots, h_{d+1}) \in L_1' \times \cdots \times L_{d+1}'$, $q$ lies in the unique bounded simplex $\Delta(h_1, \ldots, h_{d+1})$ enclosed by $\bigcup_{i=1}^{d+1} h_i$.*

---

[2]No element of $\bigcup_{i=1}^{d+1} L_i$ passes through the origin, any $d$ elements have precisely one point in common, and no $d + 1$ of them have a nonempty intersection.

*3. Tverberg-type result for simplices.* In 1998, Pach [37] showed that for all natural numbers $d$, there exists $c' = c'(d)$ with the following property. Let $P_1, P_2, \ldots, P_{d+1} \subset \mathbb{R}^d$ be disjoint $n$-element point sets with $P_1 \cup \cdots \cup P_{d+1}$ in general position. Then there is a point $q \in \mathbb{R}^d$ and subsets $P'_1 \subset P_1, \ldots, P'_{d+1} \subset P_{d+1}$, with $|P'_i| \geq c'(d)|P_i|$, such that all closed simplices with one vertex from each $P'_i$ contains $q$. The proof shows that $c'(d) = 2^{-2^{2^{O(d)}}}$. Recently Kyncl et al. [33] improved this to $c'(d) > 2^{-2^{d^2+O(d)}}$. Here we make the following improvement.

**Theorem 1.6.** *Let $P_1, P_2, \ldots, P_{d+1} \subset \mathbb{R}^d$ be disjoint $n$-element point sets with $P_1 \cup \cdots \cup P_{d+1}$ in general position. Then there is a point $q \in \mathbb{R}^d$ and subsets $P'_1 \subset P_1, \ldots, P'_{d+1} \subset P_{d+1}$, with*

$$|P'_i| \geq 2^{-O(d^2 \log(d+1))}|P_i|,$$

*such that all closed simplices with one vertex from each $P'_i$ contains $q$.*

## Algorithmic applications: Property testing.

The goal of property testing is to quickly distinguish between objects that satisfy a property from objects that are far from satisfying that property. This is an active area of computer science which was initiated by Rubinfeld and Sudan [41]. Subsequently, Goldreich, Goldwasser, and Ron [22] started the investigation of property testers for combinatorial objects.

Graph property testing, in particular, has attracted a great deal of attention. A *property* $\mathcal{P}$ is a family of graphs closed under isomorphism. A graph $G$ with $n$ vertices is $\epsilon$-*far* from satisfying $\mathcal{P}$ if one must change the adjacency relation of at least an $\epsilon$ fraction of all pairs of vertices in order to turn $G$ into a graph satisfying $\mathcal{P}$.

Let $\mathcal{F}$ be a family of graphs. An $\epsilon$-*tester* for $\mathcal{P}$ with respect to $\mathcal{F}$ is a randomized algorithm, which, given $n$ and the ability to check whether there is an edge between a given pair of vertices, distinguishes with probability at least $2/3$ between the cases $G$ satisfies $\mathcal{P}$ and $G$ is $\epsilon$-far from satisfying $\mathcal{P}$, for every $G \in \mathcal{F}$. Such an $\epsilon$-tester is *one-sided* if, whenever $G \in \mathcal{F}$ satisfies $\mathcal{P}$, the $\epsilon$-tester determines this with probability 1. A property $\mathcal{P}$ is *strongly testable* with respect to $\mathcal{F}$ if for every fixed $\epsilon > 0$, there exists a one-sided $\epsilon$-tester for $\mathcal{P}$ with respect to $\mathcal{F}$, whose query complexity is bounded only by a function of $\epsilon$, which is independent of the size of the input graph. The *vertex query complexity* of an algorithm is the number of vertices that are randomly sampled.

Property $\mathcal{P}$ is *easily testable* with respect to $\mathcal{F}$ if it is strongly testable with a one-sided $\epsilon$-tester whose query complexity is polynomial in $\epsilon^{-1}$, and otherwise $\mathcal{P}$ is *hard* with respect to $\mathcal{F}$. In classical complexity theory, an algorithm whose running time is polynomial in the input size is considered fast, and otherwise slow. This provides a nice analogue of polynomial time algorithms for property testing. The above definitions extend to $k$-uniform hypergraphs, with pairs replaced by $k$-tuples.

A very general result of Alon and Shapira [6] in graph property testing states that every hereditary family $\mathcal{P}$ of graphs is strongly testable. Unfortunately, the bounds on the query complexity that this proof gives are quite enormous. They are based on the strong regularity lemma, which gives wowzer-type bounds.[3] Even the recently improved bound by Conlon and Fox [15] still gives a tower-type bound.[4] Also, it is known that many properties are not easily testable (see [1, 2, 4]).

---

[3]Define the *tower* function $T(1) = 2$ and $T(i + 1) = 2^{T(i)}$. Then the *wowzer* function is defined as $W(1) = 2$ and $W(i + 1) = T(W(i))$.

[4]In addition to the dependence on the approximation parameter $\epsilon$, there is also a dependence on the property being tested that can make it arbitrarily hard to test [6]. However, such properties appear to be pathological and the standard properties that are studied should have only a weak dependence on the property being tested.

The result of Alon and Shapira was extended to hypergraphs by Rödl and Schacht [40]; see also the work of Austin and Tao [7]. These give even worse, Ackermann-type bounds.

In Section 5, we give an application of the polynomial semi-algebraic regularity lemma, Theorem 1.3, to show how to easily test "typical" hereditary properties with respect to semi-algebraic graphs and hypergraphs of constant complexity. The exact statement shows that the query complexity is for every property, polynomial in a natural function of that property, which one typically expects to be polynomial. This provides another example showing that semi-algebraic graphs and hypergraphs are more nicely behaved than general ones. Using the above terminology, a special case of our results is the following.

**Corollary 1.7.** *Let $H$ be a $k$-uniform hypergraph. Then $H$-freeness is easily testable within the family of semi-algebraic hypergraphs of bounded complexity.*

For background and precise results, see Section 5.

**Organization.** The rest of this paper is organized as follows. In the next section, we prove Theorem 1.1, giving a quantitative density theorem for $k$-uniform hypergraphs. In Section 3, we prove several geometric applications of this result. In Section 4, we prove the quantitative regularity lemma for semi-algebraic hypergraphs. In Section 5, we establish three results about property testing within semi-algebraic hypergraphs showing that it can be efficiently tested whether a semi-algebraic hypergraph of bounded complexity has a given hereditary property.

We systemically omit floor and ceiling signs whenever they are not crucial for the sake of clarity of presentation. We also do not make any serious attempt to optimize absolute constants in our statements and proofs.

## 2 Proof of Theorem 1.1

Let $H = (P, E)$ be a semi-algebraic $k$-uniform hypergraph in $d$-space with complexity $(t, D)$, where $P = \{p_1, \ldots, p_n\}$. Then there exists a semi-algebraic set

$$E^* = \left\{ (\mathbf{x}_1, \ldots, \mathbf{x}_k) \in \mathbb{R}^{dk} : \Phi(f_1(\mathbf{x}_1, \ldots, \mathbf{x}_k) \geq 0, \ldots, f_t(\mathbf{x}_1, \ldots, \mathbf{x}_k) \geq 0) = 1 \right\}$$

such that

$$(p_{i_1}, \ldots, p_{i_k}) \in E^* \subset \mathbb{R}^{dk} \quad \Leftrightarrow \quad (p_{i_1}, \ldots, p_{i_k}) \in E.$$

Fix $k-1$ points $q_1, \ldots, q_{k-1} \in \mathbb{R}^d$ and consider the $d$-variate polynomial $h_i(\mathbf{x}) = f_i(q_1, \ldots, q_{k-1}, \mathbf{x})$, $1 \leq i \leq t$. We use a simple but powerful trick known as *Veronese mapping (linearization)*, that transforms $h_i(\mathbf{x})$ into a linear equation. For $m = \binom{d+D}{d} - 1$, we define $\phi : \mathbb{R}^d \to \mathbb{R}^m$ to be the (Veronese) mapping given by

$$\phi(x_1, \ldots, x_d) = (x_1^{\alpha_1} \cdots x_d^{\alpha_d})_{1 \leq \alpha_1 + \cdots + \alpha_d \leq D} \in \mathbb{R}^m.$$

Then $\phi$ maps each surface $h_i(\mathbf{x}) = 0$ in $\mathbb{R}^d$ to a hyperplane $h_i^*$ in $\mathbb{R}^m$. Note that $\phi$ is injective. For example, the $d$-variate polynomial $h_i(x_1, \ldots, x_d) = c_0 + \sum_i a_i x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ would correspond to the linear equation $h_i^*(y_1, \ldots, y_m) = c_0 + \sum_{i=1}^m a_i y_i$. Given a point set $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$, we write

$$\phi(P) = \{\phi(p_1), \ldots, \phi(p_n)\} \subset \mathbb{R}^m.$$

Clearly we have the following.

**Observation 2.1.** *For $p \in \mathbb{R}^d$, we have $h_i(p) = h_i^*(\phi(p))$ and hence*

$$\mathrm{sgn}(h_i(p)) = \mathrm{sgn}(h_i^*(\phi(p))).$$

A *generalized simplex* in $\mathbb{R}^d$ is the intersection of $d+1$ half-spaces. Given a relatively open generalized simplex $\Delta \subset \mathbb{R}^m$, we say that a hyperplane $h^* \subset \mathbb{R}^m$ *crosses* $\Delta$ if it intersects $\Delta$ but does not contain it. Let us recall an old lemma due to Chazelle and Friedman (see also [14]).

**Lemma 2.2** ([13])**.** *For $m \geq 1$, let $L$ be a set of $n$ hyperplanes in $\mathbb{R}^m$, and let $r$ be an integer, $1 < r \leq n$. Then there is a subdivision of $\mathbb{R}^m$ into at most $2^{10m \log(m+1)} r^m$ relatively open generalized simplices $\Delta_i$, such that each $\Delta_i$ is crossed by at most $n/r$ hyperplanes of $L$.*

The main tool used in the proof of Theorem 1.1 is the following result on bipartite semi-algebraic graphs ($k = 2$) with point sets in different dimensions. A very similar result was obtained by Alon et al. (see Section 6 in [3]). Let $G = (P, Q, E)$ be a bipartite semi-algebraic graph, where $P \subset \mathbb{R}^{d_1}$, $Q \subset \mathbb{R}^{d_2}$, and $E \subset P \times Q$ has complexity $(t, D)$. Hence there are polynomials $f_1, f_2, \ldots, f_t$ and a boolean formula $\Phi$ such that the semi-algebraic set

$$E^* = \{(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^{d_1 + d_2} : \Phi(f_1(\mathbf{x}_1, \mathbf{x}_2) \geq 0, \ldots, f_t(\mathbf{x}_1, \mathbf{x}_2) \geq 0) = 1\},$$

satisfies

$$(p, q) \in E \quad \Leftrightarrow \quad (p, q) \in E^* \subset \mathbb{R}^{d_1 + d_2}.$$

For any point $p \in P$, the Veronese mapping $\phi$ maps the surface $f_i(p, \mathbf{x}) = 0$ in $\mathbb{R}^d$ to a hyperplane $f_i^*(p, \mathbf{y}) = 0$ in $\mathbb{R}^m$, for $1 \leq i \leq t$.

**Lemma 2.3.** *Let $G = (P, Q, E)$ be as above with $|E| \geq \varepsilon |P||Q|$. Then there are subsets $P' \subset P$ and $Q' \subset Q$ such that*

$$|P'| \geq \frac{\varepsilon}{8}|P| \quad \text{and} \quad |Q'| \geq \frac{\varepsilon^{m+1}}{t^m 2^{14m \log(m+1)}}|Q|,$$

*where $m = \binom{d_2 + D}{d_2} - 1$, and $P' \times Q' \subset E$. Moreover, there is a relatively open simplex $\Delta \subset \mathbb{R}^m$ such that $\phi(Q') = \phi(Q) \cap \Delta$, and for all $p \in P'$ and $1 \leq i \leq t$, the hyperplane $f_i^*(p, \mathbf{y}) = 0$ in $\mathbb{R}^m$ does not cross $\Delta$.*

*Proof.* Let $S$ be the set of (at most) $t|P|$ surfaces in $\mathbb{R}^{d_2}$ defined by $f_i(p, \mathbf{x}) = 0$, for $1 \leq i \leq t$ and for each $p \in P$. Set $m = \binom{d_2 + D}{d_2} - 1$ and let $\phi : \mathbb{R}^{d_2} \to \mathbb{R}^m$ denote the Veronese mapping described above. Then each surface $f_i(p, \mathbf{x}) = 0$ in $\mathbb{R}^{d_2}$ will correspond to a hyperplane $f^*(p, \mathbf{y}) = 0$ in $\mathbb{R}^m$. Then let $L$ be the set of $t|P|$ hyperplanes in $\mathbb{R}^m$ that corresponds to the surfaces in $S$.

Let $r > 1$ be an integer that will be determined later. By Lemma 2.2, there is a subdivision of $\mathbb{R}^m$ into at most $2^{10m \log(m+1)} r^m$ relatively open generalized simplices $\Delta_i$, such that each $\Delta_i$ is crossed by at most $t|P|/r$ hyperplanes from $L$. We define $Q_i \subset Q$ such that $\phi(Q_i) = \phi(Q) \cap \Delta_i$. If

$$|Q_i| < \frac{\varepsilon}{2 \cdot 2^{10m \log(m+1)} r^m}|Q|,$$

then we discard all edges in $E$ that are incident to vertices in $Q_i$. By Lemma 2.2, we have removed at most

7

$$2^{10m\log(m+1)}r^m \frac{\varepsilon}{2 \cdot 2^{10m\log(m+1)}r^m}|P||Q| = \frac{\varepsilon}{2}|P||Q|$$

edges. Let $E_1$ be the set of remaining edges. Hence, $|E_1| \geq \varepsilon|P||Q| - (\varepsilon/2)|P||Q| = (\varepsilon/2)|P||Q|$. For $p \in P$, if all $t$ hyperplanes $f_1^*(p, \mathbf{y}) = 0, \ldots, f_t^*(p, \mathbf{y}) = 0$ in $\mathbb{R}^m$ do not cross $\Delta_i$, then by Observation 2.1, the sign pattern of $f_j(p, q)$ does not change over all $q \in Q_i$. Hence in this case, we have either $p \times Q_i \subset E_1$ or $p \times Q_i \subset \overline{E}_1$.

We define $P_i \subset P$ to be the set of points in $P$ such that $p \in P_i$ if and only if $p$ is adjacent to all points in $Q_i$ (with respect to $E_1$) and all $t$ hyperplanes $f_j(p, \mathbf{x}) = 0$, $1 \leq j \leq t$, in $\mathbb{R}^m$ do not cross the simplex $\Delta_i$. By Lemma 2.2, we know that at most $t|P|/r$ points in $P$ gives rise to a hyperplane that crosses $\Delta_i$ for each $i$.

Hence there exists a $P_j$ such that $|P_j| \geq (\varepsilon/8)|P|$, since otherwise

$$|E_1| \leq \sum_i \left(\frac{\varepsilon}{8}|P| + \frac{t|P|}{r}\right)|Q_i| = \left(\frac{\varepsilon}{8} + \frac{t}{r}\right)|P||Q|.$$

For $r = 8t/\varepsilon$, this implies $|E_1| < (\varepsilon/4)|P||Q|$ which is a contradiction. Therefore we have subsets $P_j \subset P$ and $Q_j \subset Q$ such that $|P_j| \geq (\varepsilon/8)|P|$ and $P_j \times Q_j \subset E_1$. By construction of $E_1$ and for $r = 8t/\varepsilon$, we have

$$|Q_j| \geq \frac{\varepsilon^{m+1}}{t^m 2^{14m\log(m+1)}}|Q|.$$

This completes the proof.

$\square$

We are now ready to prove Theorem 1.1.

*Proof of Theorem 1.1.* Let $H = (P, E)$ be a $k$-partite semi-algebraic hypergraph in $d$-space with complexity $(t, D)$, such that $P = P_1 \cup \cdots \cup P_k$ and $|E| \geq \varepsilon|P_1| \cdots |P_k|$. For $m = \binom{d+D}{d} - 1$, we will show that there are subsets $P_1' \subset P_1, P_2' \subset P_2, \ldots, P_k' \subset P_k$, such that for $1 \leq i \leq k$,

$$|P_i'| \geq \frac{\varepsilon^{m+1}}{8^{k(m+1)}t^m(m+2)^{km}2^{14m\log(m+1)}}|P_i|,$$

and $P_1' \times \cdots \times P_k' \subset E$. This would suffice to prove the statement. We proceed by induction on $k$. The base case $k = 2$ follows from Lemma 2.3. Now assume that the statement holds for $k' < k$.

Since $E$ is semi-algebraic with complexity $(t, D)$, there are polynomials $f_1, \ldots, f_t$ and a Boolean formula $\Phi$ such that the semi-algebraic set

$$E^* = \{(\mathbf{x}_1, \ldots, \mathbf{x}_k) \in \mathbb{R}^{dk} : \Phi(f_1(\mathbf{x}_1, \ldots, \mathbf{x}_k) \geq 0, \ldots, f_t(\mathbf{x}_1, \ldots, \mathbf{x}_k) \geq 0) = 1\},$$

has the property that

$$(p_1, \ldots, p_k) \in E \quad \Leftrightarrow \quad (p_1, \ldots, p_k) \in E^* \subset \mathbb{R}^{dk}.$$

Set $Q = P_1 \times \cdots \times P_{k-1}$, and let $G = (Q, P_k, E)$ be the bipartite semi-algebraic graph with relation $E$. Just as above, we let $\phi : \mathbb{R}^d \to \mathbb{R}^m$ denote the Veronese mapping. Then for $(p_1, \ldots, p_{k-1}) \in Q$, each surface $f_i(p_1, \ldots, p_{k-1}, \mathbf{x}) = 0$ in $\mathbb{R}^d$ will correspond to a hyperplane $f_i^*(p_1, \ldots, p_{k-1}, \mathbf{y}) = 0$ in $\mathbb{R}^m$, $1 \leq i \leq t$. By Lemma 2.3, there are subsets $Q' \subset Q$ and $P_k' \subset P_k$ such that

$$|Q'| \geq \frac{\varepsilon}{8}|Q| \quad \text{and} \quad |P_k'| \geq \frac{\varepsilon^{m+1}}{t^m 2^{14m \log(m+1)}}|P_k|,$$

and $Q' \times P_k' \subset E$. Moreover there is a relatively open simplex $\Delta \subset \mathbb{R}^m$ such that $\phi(P_k') = \Delta \cap \phi(P_k)$, and for any $(k-1)$-tuple $(p_1, \ldots, p_{k-1}) \in Q'$, all $t$ hyperplanes $f_i^*(p_1, \ldots, p_{k-1}, \mathbf{y}) = 0$, $1 \leq i \leq t$, do not cross $\Delta$. Let $p_1^*, \ldots, p_\ell^* \in \mathbb{R}^m$ be the vertices of $\Delta$, where $\ell \leq m+1$.

We define $H_1 = (P \setminus P_k, E_1)$ to be the $(k-1)$-partite semi-algebraic hypergraph in $d$-space on the point set $P_1 \cup \cdots \cup P_{k-1}$ with relation $E_1 \subset P_1 \times \cdots \times P_{k-1}$, where $(p_1, \ldots, p_{k-1}) \in E_1$ if and only if $(p_1, \ldots, p_{k-1}, q) \in E$ for all $q \in P_k'$, and all $t$ hyperplanes $f_i^*(p_1, \ldots, p_{k-1}, \mathbf{y}) = 0$ in $\mathbb{R}^m$, $1 \leq i \leq t$, do not cross $\Delta$.

Next we need to check that the complexity of $E_1$ is not too high. In order to do this, we define relations $E_2, E_3 \subset P_1 \times \cdots \times P_{k-1}$ as follows. Fix a point $q \in P_k'$, and let

$$E_2 = \{(p_1, \ldots, p_{k-1}) \in P_1 \times \cdots \times P_{k-1} : (p_1, \ldots, p_{k-1}, q) \in E\}.$$

Since the relation $E$ has complexity $(t, D)$, the relation $E_2$ is semi-algebraic with complexity $(t, D)$.

We define the relation $E_3 \subset P_1 \times \cdots \times P_{k-1}$, where $(p_1, \ldots, p_{k-1}) \in E_3$ if and only if all $t$ hyperplanes $f_i^*(p_1, \ldots, p_{k-1}, \mathbf{y}) = 0$, $1 \leq i \leq t$, do not cross $\Delta$ in $\mathbb{R}^m$. Notice that a hyperplane does not cross $\Delta$ in $\mathbb{R}^m$ if and only if the vertices $p_1^*, \ldots, p_\ell^*$ of $\Delta$ all lie in a closed half-space defined by the hyperplane (recall that $\Delta$ is relatively open). Therefore, for each vertex $p_j^*$ of the simplex $\Delta$, we define $t$ $(dk - d)$-variate polynomials

$$g_{i,j}(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}) = f_i^*(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, p_j^*).$$

Then there is a Boolean formula $\Phi_3$ such that the semi-algebraic set

$$E_3^* = \{(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}) \in \mathbb{R}^{d(k-1)} : \Phi_3(g_{1,1}(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}) \geq 0, \ldots, g_{t,\ell}(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}) \geq 0)\}$$

has the property that $(p_1, \ldots, p_{k-1}) \in E_3$ if and only if $(p_1, \ldots, p_{k-1}) \in E_3^* \subset \mathbb{R}^{d(k-1)}$, and $E_3$ has complexity $(t\ell, D)$. Notice that if $(p_1, \ldots, p_{k-1}) \in E_3$, then by Observation 2.1, the sign pattern of $f_i(p_1, \ldots, p_{k-1}, q)$ does not change over all $q \in P_k'$ since $\phi(P_k') \subset \Delta$. Therefore, if we also have $(p_1, \ldots, p_{k-1}) \in E_2$, this implies $(p_1, \ldots, p_{k-1}) \in E_1$. That is,

$$(p_1, \ldots, p_{k-1}) \in E_1 \quad \Leftrightarrow \quad (p_1, \ldots, p_{k-1}) \in E_1^* = E_2^* \cap E_3^* \subset \mathbb{R}^{d(k-1)},$$

and $E_1$ has complexity $(t + t\ell, D)$. Since $\ell \leq m+1$, $E_1$ has complexity $(t(m+2), D)$. By construction of $Q'$, we have

$$|E_1| \geq |Q'| \geq \frac{\varepsilon}{8}|P_1| \cdots |P_{k-1}|.$$

By applying induction hypothesis on the $(k-1)$-partite semi-algebraic hypergraph $H_1 = (P \setminus P_k, E_1)$, we obtain subsets $P_i' \subset P_i$, $1 \leq i \leq k-1$, such that

$$\begin{aligned} |P_i'| &\geq \frac{(\varepsilon/8)^{m+1}}{8^{(k-1)(m+1)}(t(m+2))^m(m+2)^{(k-1)m}2^{14m \log(m+1)}}|P_i| \\ &\geq \frac{\varepsilon^{m+1}}{8^{k(m+1)}t^m(m+2)^{km}2^{14m \log(m+1)}}|P_i|, \end{aligned}$$

and $P_1' \times \cdots \times P_{k-1}' \subset E_1$. By definition of $E_1$ and by construction of $P_k'$, we have $P_1' \times \cdots \times P_k' \subset E$, and this completes the proof. $\qquad\square$

# 3 Applications of Corollary 1.2

In this section, we give three applications of Corollary 1.2.

## 3.1 Same-type lemma

*Proof of Theorem 1.4.* Let $P_1, \ldots, P_k$ be finite point sets in $\mathbb{R}^d$ such that $P = P_1 \cup \cdots \cup P_k$ is in general position. By a result of Goodman and Pollack (see [24] and [23]), the number of different order-types of $k$-element point sets in $d$ dimensions is at most $k^{O(d^2 k)}$. By the pigeonhole principle, there is an order-type $\pi$ such that at least

$$k^{-O(d^2 k)} |P_1| \cdots |P_k|$$

$k$-tuples $(p_1, \ldots, p_k) \in (P_1, \ldots, P_k)$ have order-type $\pi$. We define the relation $E \subset P_1 \times \cdots \times P_k$, where $(p_1, \ldots, p_k) \in E$ if and only if $(p_1, \ldots, p_k)$ has order-type $\pi$. Next we need to check that the complexity of $E$ is not too high.

We can check to see if $(p_1, \ldots, p_k)$ has order-type $\pi$ by simply checking the orientation of every $(d+1)$-tuple of $(p_1, \ldots, p_k)$. More specifically, for $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,d})$, we define the $(d^2 + d)$-variate polynomial

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_{d+1}) = \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{d+1,1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,d} & x_{2,d} & \cdots & x_{d+1,d} \end{pmatrix}.$$

Then there exists a Boolean formula $\Phi$, such that the semi-algebraic set

$$E^* = \{(\mathbf{x}_1, \ldots, \mathbf{x}_k) \in \mathbb{R}^{dk} : \Phi(\{f(\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{d+1}}) \geq 0\}_{1 \leq i_1 < \cdots < i_{d+1} \leq k}) = 1\},$$

which is defined by the $\binom{k}{d+1}$ polynomials $f(\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{d+1}})$, $1 \leq i_1 < \cdots < i_{d+1} \leq k$, has the property that

$$(p_1, \ldots, p_k) \in E \quad \Leftrightarrow \quad (p_1, \ldots, p_k) \in E^*.$$

Hence the complexity of $E$ is $\left(\binom{k}{d+1}, 1\right)$. Therefore we can apply Corollary 1.2 to the $k$-partite semi-algebraic hypergraph $H = (P, E)$ with $\varepsilon = k^{-O(d^2 k)}$ and $t = \binom{k}{d+1}$ to obtain subsets $P_1' \subset P_1, \ldots, P_k' \subset P_k$ such that $(P_1', \ldots, P_k')$ has same-type transversals (each transversal has order-type $\pi$), and

$$|P_i'| \geq 2^{-O(d^3 k \log k)} |P_i|,$$

for $1 \leq i \leq k$. $\qquad\square$

## 3.2 A Tverberg-type result

The proof of Theorem 1.6 requires the following due to Karasev [30] (see also Theorem 4 in [33]).

**Lemma 3.1** ([30]). *Let $P_1, \ldots, P_{d+1} \subset \mathbb{R}^d$ be disjoint $n$-element point sets with $P_1 \cup \cdots \cup P_{d+1}$ in general position. Then there is a point $q \in \mathbb{R}^d$ which is contained in the interior of at least*

$$\frac{1}{(d+1)!}n^{d+1}$$

*rainbow simplices, where a rainbow simplex is a simplex generated by selecting one point from each $P_i$.*

*Proof of Theorem 1.6.* We may assume $n = 2^{\Omega(d^2 \log d)}$ as otherwise we can take $p_i \in P_i$, $P_i' = \{p_i\}$, and $q$ to be any point in the simplex with vertices $p_1, \ldots, p_{d+1}$.

Let $P_1, \ldots, P_{d+1}$ be disjoint $n$-element point sets with $P_1 \cup \cdots \cup P_{d+1}$ in general position. If a simplex has one vertex in each $P_i$, then we call it *rainbow*. Hence the number of rainbow simplices is $N = n^{d+1}$.

By Lemma 3.1, there is a point $q$ contained in the interior of at least

$$\frac{1}{(d+1)!}n^{d+1}$$

rainbow simplices.

We define $H = (P, E)$ to be the $(d+1)$-partite semi-algebraic hypergraph, where $P = P_1 \cup \cdots \cup P_{d+1}$, $E \subset P_1 \times \cdots \times P_{d+1}$, where $(p_1, \ldots, p_{d+1}) \in E$ if and only if $q \in \text{conv}(p_1 \cup \cdots \cup p_{d+1})$. Next we need to check that the complexity of $E$ is not too high.

To see if $q \in \text{conv}(p_1 \cup \cdots \cup p_{d+1})$, we just need to check that the points $q$ and $p_j$ lie on the same side of the hyperplane spanned by the $d$-tuple $(p_1, \ldots, p_{j-1}, p_{j+1}, \ldots, p_{d+1})$, for each $1 \leq j \leq d+1$. More specifically, for $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,d})$, we define the $(d^2 + d)$-variate polynomial

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_{d+1}) = \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{1,1} & x_{2,1} & \cdots & x_{d+1,1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1,d} & x_{2,d} & \cdots & x_{d+1,d} \end{pmatrix}.$$

Then there exists a Boolean formula $\Phi$ such that the semi-algebraic set

$$E^* = \left\{ (\mathbf{x}_1, \ldots, \mathbf{x}_{d+1}) \in \mathbb{R}^{d(d+1)} : \Phi \left( \left\{ \begin{array}{l} f(\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_{d+1}, q) \geq 0, \\ f(\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_{d+1}, \mathbf{x}_j) \geq 0 \end{array} \right\}_{1 \leq j \leq d+1} \right) = 1 \right\},$$

satisfies

$$(p_1, \ldots, p_{d+1}) \in E \quad \Leftrightarrow \quad (p_1, \ldots, p_{d+1}) \in E^*.$$

Hence $E$ has complexity $(2(d+1), 1)$, and

$$|E| \geq \frac{1}{(d+1)!}|P_1| \cdots |P_{d+1}|.$$

11

Hence we can apply Corollary 1.2 to $H = (P, E)$ with $\varepsilon = 1/(d+1)!$ and $t = 2(d+1)$, to obtain subsets $P_1' \subset P_1, \dots, P_{d+1}' \subset P_{d+1}$ such that

$$|P_i'| \geq 2^{-O(d^2 \log(d+1))} |P_i|,$$

such that all closed simplices with one vertex from each $P_i'$ contains $q$. $\qquad\square$

# 4  Regularity lemma for semi-algebraic hypergraphs

In this section, we will prove Theorem 4.2. First we recall the definition of a semi-algebraic set in $\mathbb{R}^d$. A set $A \subset \mathbb{R}^d$ is *semi-algebraic* if there are polynomials $f_1, f_2, \dots, f_t \in \mathbb{R}[x_1, \dots, x_d]$ and a Boolean function $\Phi$ such that

$$A = \left\{ \mathbf{x} \in \mathbb{R}^d : \Phi(f_1(\mathbf{x}) \geq 0, \dots, f_t(\mathbf{x}) \geq 0) = 1 \right\}.$$

We say that a semi-algebraic set in $d$-space has *description complexity at most* $\kappa$ if the number of inequalities is at most $\kappa$, and each polynomial $f_i$ has degree at most $\kappa$. By setting $\varepsilon = 1/2$ in Theorem 1.1, we obtain the following Ramsey-type result due to Fox, Gromov *et al.*, see also Bukh and Hubard for another proof.

**Lemma 4.1** (Theorem 18 in [11]). *Let $H = (P, E)$ be a $k$-partite semi-algebraic hypergraph in $d$-space, where $P = P_1 \cup \cdots \cup P_k$, $E \subset P_1 \times \cdots \times P_k$, and $E$ has complexity $(t, D)$. Then there exists a $\delta = \delta(k, d, t, D)$ and subsets $P_1' \subset P_1, \dots, P_k' \subset P_k$ such that for $1 \leq i \leq k$,*

$$|P_i'| \geq \delta |P_i|,$$

*and $(P_1', \dots, P_k')$ is homogeneous, i.e., either complete or empty. Moreover, there are semi-algebraic sets $\Delta_1, \dots, \Delta_k \subset \mathbb{R}^d$ such that $\Delta_i$ has complexity $\kappa$, where $\kappa = \kappa(d, t, D)$, and $P_i' = P_i \cap \Delta_i$ for all $i$.*

While the above lemma in [19] and [11] does not explicitly state that the constructed $P_i'$ is the intersection of $P_i$ with a semi-algebraic set of low complexity, the proof is easily seen to give this.

We note that Theorem 18 in [11] and Theorem 8.1 in [19] does not include the statement that there are semi-algebraic sets $\Delta_1, \dots, \Delta_k \subset \mathbb{R}^d$ such that $P_i' = P_i \cap \Delta_i$ for all $i$. However their proof, as well as the proof of Theorem 1.1 in Section 2, shows that such semi-algebraic sets exist.

We will first prove the following variant of Theorem 1.3, from which Theorem 1.3 quickly follows.

**Theorem 4.2.** *For any $\varepsilon > 0$, the vertex set of any semi-algebraic $k$-uniform hypergraph $H = (P, E)$ in $d$-space with complexity $(t, D)$, can be partitioned into $K \leq (1/\varepsilon)^c$ parts $P = P_1 \cup \cdots \cup P_K$, where $c = c(k, d, t, D)$, such that*

$$\sum \frac{|P_{j_1}| \cdots |P_{j_k}|}{|P|^k} \leq \varepsilon,$$

*where the sum is taken over all $k$-tuples $(j_1, \dots, j_k)$ such that $(P_{j_1}, \dots, P_{j_k})$ is not homogeneous.*

*Proof.* Let $\varepsilon > 0$ and $H = (P, E)$ be an $n$-vertex $k$-uniform semi-algebraic hypergraph with complexity $(t, D)$ in $d$-space. For integer $r \geq 0$, we will recursively define a partition $\mathcal{P}_r$ on $P^k = P \times \cdots \times P$ into at most $2^{kr}$ parts of the form $X_1 \times \cdots \times X_k$, such that at most $(1 - \delta^k)^r |P|^k$

12

$k$-tuples $(p_1, \ldots, p_k)$ lie in a part $X_1 \times \cdots \times X_k$ with the property that $(X_1, \ldots, X_k)$ is not homogeneous. Note that $\delta$ is defined in Lemma 4.1. Also for each $r \geq 0$, we will inductively define a collection $\mathcal{F}_r$ of semi-algebraic sets, each set with complexity at most $c = c(d, t, D)$, such that $|\mathcal{F}_r| \leq \sum_{j=0}^{r} 2^{kj}$, and for each part $X_1 \times \cdots \times X_k$ in $\mathcal{P}_r$, there are subcollections $\mathcal{S}_1, \ldots, \mathcal{S}_k \subset \mathcal{F}_r$ such that

$$X_i = \left( \bigcap_{\Delta \in \mathcal{S}_i} \Delta \right) \cap P.$$

Note that $c = c(d, t, D)$ is defined in Lemma 4.1. Given such a partition $\mathcal{P}_r$, a $k$-tuple $(p_1, \ldots, p_k) \in P \times \cdots \times P$ is called *bad* if $(p_1, \ldots, p_k)$ lies in a part $X_1 \times \cdots \times X_k$ in $\mathcal{P}_r$, for which $(X_1, \ldots, X_k)$ is not homogeneous.

We start with $\mathcal{P}_0 = \{P \times \cdots \times P\}$ and $\mathcal{F}_0 = \{\mathbb{R}^d\}$, which satisfies the base case $r = 0$. After obtaining $\mathcal{P}_i$ and $\mathcal{F}_i$, we define $\mathcal{P}_{i+1}$ and $\mathcal{F}_{i+1}$ as follows. Let $X_1 \times \cdots \times X_k$ be a part in the partition $\mathcal{P}_i$. Then if $(X_1, \ldots, X_k)$ is homogeneous, we put $X_1 \times \cdots \times X_k$ in $\mathcal{P}_{i+1}$. If $(X_1, \ldots, X_k)$ is not homogeneous, then notice that $(X_1, \ldots, X_k)$ gives rise to $|X_1| \cdots |X_k|$ bad $k$-tuples $(p_1, \ldots, p_k)$. Hence we apply Lemma 4.1 on $(X_1, \ldots, X_k)$ to obtain subsets $X_1' \subset X_1, \ldots, X_k' \subset X_k$ and semi-algebraic sets $\Delta_1, \ldots, \Delta_k$ with the properties described above. Then we partition $X_1 \times \cdots \times X_k$ into $2^k$ parts $Z_1 \times \cdots \times Z_k$ where $Z_i \in \{X_i', X_i \setminus X_i'\}$ for $1 \leq i \leq k$, and put these parts into $\mathcal{P}_{i+1}$. The collection $\mathcal{F}_{i+1}$ will consist of all semi-algebraic sets from $\mathcal{F}_i$, plus all semi-algebraic sets $\Delta_1, \ldots, \Delta_k, \mathbb{R}^d \setminus \Delta_1, \ldots, \mathbb{R}^d \setminus \Delta_k$ that was obtained after applying Lemma 4.1 to each part $X_1 \times \cdots \times X_k$ in $\mathcal{P}_r$ for which $(X_1, \ldots, X_k)$ is not homogeneous.

By the induction hypothesis, the number of bad $k$-tuples $(p_1, \ldots, p_k)$ in $\mathcal{P}_{i+1}$ is at most

$$(1 - \delta^k)(1 - \delta^k)^i |P|^k = (1 - \delta^k)^{i+1} |P|^k.$$

The number of parts in $\mathcal{P}_{i+1}$ is at most $2^k \cdot 2^{ki} = 2^{k(i+1)}$, and $|\mathcal{F}_{i+1}| \leq |\mathcal{F}_i| + 2k|\mathcal{P}_i| \leq |\mathcal{F}_i| + 2^{k(i+1)} \leq \sum_{j=0}^{i+1} 2^{kj}$. By the induction hypothesis, for any part $X_1 \times \cdots \times X_k$ in $\mathcal{P}_{i+1}$ such that $X_1 \times \cdots \times X_k$ was also in $\mathcal{P}_i$, there are subcollections $\mathcal{S}_1, \ldots, \mathcal{S}_k \subset \mathcal{F}_i$ such that

$$X_i = \left( \bigcap_{\Delta \in \mathcal{S}_i} \Delta \right) \cap P,$$

for $1 \leq i \leq k$. If $X_1 \times \cdots \times X_k$ is not in $\mathcal{P}_i$, then there must be a part $Y_1 \times \cdots \times Y_k$ in $\mathcal{P}_i$, such that $X_1 \times \cdots \times X_k$ is one of the $2^k$ parts obtained from applying Lemma 4.1 to $Y_1 \times \cdots \times Y_k$. Hence, $X_i \subset Y_i$ for $1 \leq i \leq k$. Let $\Delta_1, \ldots, \Delta_k$ be the semi-algebraic sets obtained when applying Lemma 4.1 to $Y_1 \times \cdots \times Y_k$. By the induction hypothesis, we know that there are subcollections $\mathcal{S}_1, \ldots, \mathcal{S}_k \subset \mathcal{F}_i$ such that

$$Y_i = \left( \bigcap_{\Delta \in \mathcal{S}_i} \Delta \right) \cap P,$$

for $1 \leq i \leq k$. Hence, there are subcollections $\mathcal{S}_i' \subset \mathcal{S}_i \cup \{\Delta_i, \mathbb{R}^d \setminus \Delta_i\}$ such that

$$X_i = \left( \bigcap_{\Delta \in \mathcal{S}_i'} \Delta \right) \cap P,$$

13

for $1 \le i \le k$. We have therefore obtained our desired partition $\mathcal{P}_{i+1}$ on $P \times \cdots \times P$, and collection $\mathcal{F}_{i+1}$ of semi-algebraic sets.

At step $r = \frac{\log \varepsilon}{\log(1-\delta^k)}$, there are at most $\varepsilon |P|^k$ bad $k$-tuples $(p_1, \ldots, p_k)$ in partition $\mathcal{P}_r$. The number of parts of $\mathcal{P}_r$ is at most $(1/\varepsilon)^{c_1}$ and $|\mathcal{F}_r| \le (1/\varepsilon)^{c_2}$, where $c_1 = c_1(k,d,t,D)$ and $c_2 = c_2(k,d,t,D)$ (recall that $\delta = \delta(k,d,t,D)$).

Finally, we partition the vertex set $P$ into $K$ parts, $P_1, P_2, \ldots, P_K$, such that two vertices are in the same part if and only if every member of $\mathcal{F}_r$ contains both or neither of them. Since $\mathcal{F}_r$ consists of at most $(1/\varepsilon)^{c_2}$ semi-algebraic sets, and each set has complexity at most $c = c(k,d,t,D)$, we have $K \le (1/\varepsilon)^{c_3}$ where $c_3 = c_3(k,d,t,D)$ (see Theorem 6.2.1 in [34]). Now we just need to show that

$$\sum |P_{j_1}| \cdots |P_{j_k}| < \varepsilon |P|^k,$$

where the sum is taken over all $k$-tuples $(j_1, \ldots, j_k)$, $1 \le j_1 < \cdots < j_k \le K$, such that $(P_{j_1}, \ldots, P_{j_k})$ is not homogeneous. It suffices to show that for a $k$-tuple $(j_1, \ldots, j_k)$, if $(P_{j_1}, \ldots, P_{j_k})$ is not homogeneous, then all $k$-tuples $(p_1, \ldots, p_k) \in P_{j_1} \times \cdots \times P_{j_k}$ are bad in the partition $\mathcal{P}_r$.

For the sake of contradiction, suppose $(P_{j_1}, \ldots, P_{j_k})$ is not homogeneous, and the $k$-tuple $(p_1, \ldots, p_k) \in P_{j_1} \times \cdots \times P_{j_k}$ is not a bad $k$-tuple. Then there is a part $X_1 \times \cdots \times X_k$ in the partition $\mathcal{P}_r$ such that $p_i \in X_i$ for all $i$, and $(X_1, \ldots, X_k)$ is homogeneous. Hence there are subcollections $\mathcal{S}_1, \ldots, \mathcal{S}_k \subset \mathcal{F}_r$ such that

$$X_i = \left( \bigcap_{\Delta \in \mathcal{S}_i} \Delta \right) \cap P,$$

for $1 \le i \le k$. However, by construction of $P_{j_1}, \ldots, P_{j_k}$, this implies that $P_{j_i} \subset X_i$ for all $i$, and we have a contradiction. Therefore, we have obtained our desired partition $P_1, P_2, \ldots, P_K$. $\square$

*Proof of Theorem 1.3.* Apply Theorem 4.2 with approximation parameter $\varepsilon/2$. So there is a partition $\mathcal{Q} : P = Q_1 \cup \cdots \cup Q_{K'}$ into $K' \le (2/\varepsilon)^c$ parts, where $c = c(k,d,t,D)$, such that $\sum |Q_{i_1}||Q_{i_2}| \cdots |Q_{i_k}| \le (\varepsilon/2)|P|^k$, where the sum is taken over all $k$-tuples $(i_1, \ldots, i_k)$ such that $(Q_{i_1}, \ldots, Q_{i_k})$ is not homogeneous.

Let $K = 4k\varepsilon^{-1}K'$. Partition each part $Q_i$ into parts of size $|P|/K$ and possibly one additional part of size less than $|P|/K$. Collect these additional parts and divide them into parts of size $|P|/K$ to obtain an equitable partition $\mathcal{P} : P = P_1 \cup \cdots \cup P_K$ into $K$ parts. The number of vertices of $P$ which are in parts $P_i$ that are not contained in a part of $\mathcal{Q}$ is at most $K'|P|/K$. Hence, the fraction of $k$-tuples $P_{i_1} \times \cdots \times P_{i_k}$ with not all $P_{i_1}, \ldots, P_{i_k}$ subsets of parts of $\mathcal{Q}$ is at most $kK'/K = \varepsilon/4$. As $\varepsilon/2 + \varepsilon/4 < \varepsilon$, we obtain that less than an $\varepsilon$-fraction of the $k$-tuples of parts of $\mathcal{P}$ are not homogeneous, which completes the proof. $\square$

# 5 Property testing in semi-algebraic hypergraphs

In this section, we apply the polynomial semi-algebraic regularity lemma, Theorem 1.3, to quickly distinguish between semi-algebraic objects that satisfy a property from objects that are far from satisfying it. In the first subsection, we restrict ourselves to testing *monotone* hypergraph properties. We then discuss and prove a result about easily testing *hereditary* properties of graphs. We conclude with a result on easily testing hypergraph hereditary properties. All hypergraphs we consider in this section are assumed to be $k$-uniform.

## 5.1 Testing monotone properties

Let $\mathcal{P}$ be a monotone property of hypergraphs, and $\mathcal{H}$ be the family of minimal forbidden hypergraphs for $\mathcal{P}$. That is, $H \in \mathcal{H}$ if $H \notin \mathcal{P}$, but every proper subhypergraph of $H$ is in $\mathcal{P}$. We say that a hypergraph $H$ has a *homomorphism* to another hypergraph $R$, and write $H \to R$, if there is a mapping $f : V(H) \to V(R)$ such that the image of every edge of $H$ is an edge of $R$.

We let $\mathcal{H}_r$ denote the family of hypergraphs $R$ on at most $r$ vertices for which there is a hypergraph $H \in \mathcal{H}$ with $H \to R$. Define

$$\Psi_1(\mathcal{H}, r) = \max_{R \in \mathcal{H}_r} \min_{H \in \mathcal{H}, H \to R} |V(H)|.$$

The following result implies that we can easily test every monotone property $\mathcal{P}$ whose corresponding function $\Psi_1(\mathcal{H}, r)$ grows at most polynomially in $r$. A simple example in which $\Psi_1(\mathcal{H}, r)$ is constant is the case that the property $\mathcal{P}$ is $H$-freeness for a fixed hypergraph $H$, i.e., $\mathcal{P}$ is the family of $k$-uniform hypergraphs which do not contain $H$ as a subhypergraph.

**Theorem 5.1.** *Let $\mathcal{P}$ be a monotone property of hypergraphs, and $\mathcal{H}$ be the family of minimal forbidden hypergraphs for $\mathcal{P}$. Within the family $\mathcal{A}$ of semi-algebraic hypergraphs in $d$-space with description complexity $(t, D)$, the property $\mathcal{P}$ can be $\epsilon$-tested with vertex query complexity at most $8r\Psi_1(\mathcal{H}, r)$, where $r = (1/\epsilon)^c$ with $c = c(k, d, t, D)$ is the number of parts in the algebraic regularity lemma for hypergraphs as in Theorem 1.3.*

*Proof.* Let $s = \Psi_1(\mathcal{H}, r)$ and $v = 8rs$. Consider the tester which samples $v$ vertices from a hypergraph $A \in \mathcal{A}$. It accepts if the induced subhypergraph on these $v$ vertices has property $\mathcal{P}$ and rejects otherwise. It suffices to show that if $A \in \mathcal{A}$ is $\epsilon$-far from satisfying $\mathcal{P}$, then with probability at least $2/3$, the tester will reject.

Consider an equitable partition $\mathcal{Q} : V(A) = V_1 \cup \ldots \cup V_{r'}$ of the vertex set of $A$ guaranteed by Theorem 1.3 such that all but at most an $\epsilon$-fraction of the $k$-tuples of parts are homogeneous. Let $r' = |\mathcal{Q}|$ be the number of parts of the partition, so $r' \leq r$. Delete all edges of $A$ whose vertices go between parts which are not complete. By the almost homogeneous property of the partition, at most an $\epsilon$-fraction of the edges are deleted. Let $A'$ denote the resulting subhypergraph of $A$. As $\mathcal{P}$ is monotone, if $A \in \mathcal{P}$, then $A' \in \mathcal{P}$. Let $R$ be the hypergraph on $[r']$, one for each part of $\mathcal{Q}$, and a $k$-tuple $(i_1, \ldots, i_k)$ of (not necessarily distinct) vertices of $R$ forms an edge if and only if the corresponding $k$-tuple $V_{i_1}, \ldots, V_{i_k}$ of parts are complete in $A'$ (and hence in $A$ as well). If $R \notin \mathcal{H}_r$, then every hypergraph which has a homomorphism to $R$ is in $\mathcal{P}$ and hence $A' \in \mathcal{P}$. However, at most an $\epsilon$-fraction of the $k$-tuples are deleted from $A$ to obtain $A'$, and so $A$ is not $\epsilon$-far from satisfying $\mathcal{P}$, contradicting the assumption. Hence, $R \in \mathcal{H}_r$. Since $R \in \mathcal{H}_r$, there is a hypergraph $H \in \mathcal{H}$ on at most $s = \Psi_1(\mathcal{H}, r)$ vertices with $H \to R$. Consider such a homomorphism $f : V(H) \to V(R)$, and let $a_i = |f^{-1}(i)|$. If the sampled $v$ vertices contains at least $a_i$ vertices in $V_i$ for each $i$, then $H$ is a subgraph of the sampled vertices and hence the sampled vertices does not have property $\mathcal{P}$.

So we need to estimate the probability of the event that the sampled $v$ vertices contains $a_i$ vertices in $V_i$ for each $i$. For a particular $i$, the probability that the sampled $v$ vertices contains

15

fewer than $a = a_i$ vertices in $V_i$ is 0 if $a = 0$ and is otherwise

$$\sum_{j=0}^{a-1}\binom{|V_i|}{j}\binom{|V(A)|-|V_i|}{v-j}\Big/\binom{|V(A)|}{v} \;<\; (1/r')^a(1-1/r')^{v-a}\binom{v}{a} = (r'-1)^{-a}(1-1/r')^v\binom{v}{a}$$

$$< \; (r'-1)^{-a}e^{-v/r'}(ve/a)^a < e^{-8s(r/r')}(30(r/r')s/a)^a$$
$$\le \; e^{-8s(r/r')}(30(r/r'))^s \le e^{-8s}30^s < 1/(4s),$$

where in the last inequality we use that $a \le s$. Taking the union bound and summing over all $i$, the probability that the sampled set of $v$ vertices does not contain $a_i$ vertices in $V_i$ for at least one $i$ is at most $s \times 1/(4s) = 1/4$. This completes the proof. $\square$

## 5.2 Testing hereditary properties of graphs

We next state and prove a result which shows that typical hereditary properties of graphs are easily testable within semi-algebraic graphs. We say that for a graph $H$ and a graph $R$ on $[r']$ with loops, there is an *induced homomorphism* from $H$ to $R$, and we write $H \to_{\text{ind}} R$, if there is a mapping $f : V(H) \to V(R)$ which maps edges of $H$ to edges of $R$, and every nonadjacent pair of distinct vertices of $H$ gets mapped to a nonadjacent pair in $R$. We write $H \not\to_{\text{ind}} R$ if $H \to_{\text{ind}} R$ does not hold.

Let $P = P_1 \cup \ldots \cup P_{r'}$ be a vertex partition of a semi-algebraic graph $G$. A key observation is that if we *round* $G$ by the partition of $P$ and the graph $R$ with loops to obtain a graph $G'$ on the same vertex set as $G$ by adding edges to make $P_i, P_j$ complete if $(i,j)$ is an edge of $R$, and deleting edges to make $P_i, P_j$ empty if $(i,j)$ is not an edge of $R$ and we have that $H \not\to_{\text{ind}} R$, then $G'$ does not contain $H$ as an induced subgraph.

Let $\mathcal{P}$ be a hereditary graph property, and $\mathcal{H}$ be the family of minimal (induced) forbidden graphs for $\mathcal{P}$. That is, each $H \in \mathcal{H}$ satisfies $H \notin \mathcal{P}$, but every proper induced subgraph $H'$ of $H$ satisfies $H' \in \mathcal{P}$. For a nonnegative integer $r$, let $\mathcal{H}_r$ be the family of graphs $R$ on at most $r$ vertices for which there is at least one $H \in \mathcal{H}$ such that $H \to_{\text{ind}} R$. As long as $\mathcal{H}_r$ is nonempty, define

$$\Psi_2(\mathcal{H}, r) = \max_{R \in \mathcal{H}_r} \min_{H \in \mathcal{H}: H \to_{\text{ind}} R} |V(H)|.$$

If $\mathcal{H}_r$ is empty, then we define $\Psi_2(\mathcal{H}, r) = 1$. Note that $\Psi_2(\mathcal{H}, r)$ is a monotonically increasing function of $r$.

We now state our main result for testing hereditary graph properties within semi-algebraic graphs. It implies that, if $\Psi_2(\mathcal{H}, r)$ is at most polynomial in $r$, then $\mathcal{P}$ can be easily tested within the family of semi-algebraic graphs of constant description complexity, i.e., there is an $\epsilon$-tester with vertex query complexity $\epsilon^{-O(1)}$. A simple example for which $\Psi_2(\mathcal{H}, r)$ is constant is the case that the property $\mathcal{P}$ is the family of graphs which do not contain an induced subgraph isomorphic to $H$, for some fixed graph $H$.

**Theorem 5.2.** *Let $\mathcal{P}$ be a hereditary property of graphs, and $\mathcal{H}$ be the family of minimal forbidden graphs for $\mathcal{P}$. Within the family $\mathcal{A}$ of semi-algebraic graphs in d-space with description complexity $(t, D)$, property $\mathcal{P}$ can be $\epsilon$-tested with vertex query complexity at most $(r\Psi_2(\mathcal{H}, r))^C$, where $r = (1/\epsilon)^C$ with $C = C(d, t, D)$.*

We show how this theorem can be established using the following "strong regularity lemma" for semi-algebraic graphs.

**Theorem 5.3.** *For any $0 < \alpha, \epsilon < 1/2$, any semi-algebraic graph $H = (P, E)$ in d-space with complexity $(t, D)$ has an equitable vertex partition $P = P_1 \cup \cdots \cup P_{r'}$ with $r' \leq r = (1/\epsilon)^{c'}$ with $c' = c'(d, t, D)$ such that all but an $\epsilon$-fraction of the pairs $P_i, P_j$ are homogeneous. Furthermore, there are subsets $Q_i \subset P_i$ such that each pair $Q_i, Q_j$ with $i \neq j$ is complete or empty, and each $Q_i$ has density at most $\alpha$ or at least $1 - \alpha$. Moreover, $|Q_i| \geq \delta|P|$ with $\delta = (\alpha\epsilon)^{-c}$ with $c = c(d, t, D)$.*

We next prove Theorem 5.2 assuming Theorem 5.3. The rest of the subsection is then devoted to proving Theorem 5.3.

*Proof of Theorem 5.2.* Let $s = \Psi_2(\mathcal{H}, r)$ and $v = (rs)^C$ for an appropriate constant $C = C(d, t, D)$. Consider the tester which samples $v$ vertices from a graph $A = (P, E) \in \mathcal{A}$. It accepts if the induced subgraph on these $v$ vertices has property $\mathcal{P}$ and rejects otherwise. It suffices to show that if $A \in \mathcal{A}$ is $\epsilon$-far from satisfying $\mathcal{P}$, then with probability at least $2/3$, the tester will reject.

Consider an equitable partition $P = P_1 \cup \ldots \cup P_{r'}$ with $r' \leq r = (1/\epsilon)^{c'}$ of the vertex set of $A$ guaranteed by Theorem 5.3 with the property that all but at most an $\epsilon$-fraction of the pairs of parts are homogeneous, and, with $\alpha = 1/(4s^2)$, there are subsets $Q_i \subset P_i$ such that each pair $Q_i, Q_j$ with $i \neq j$ is complete or empty, each $Q_i$ has density at most $\alpha$ or at least $1 - \alpha$, and $|Q_i| \geq \delta|P|$ with $\delta = (\alpha\epsilon)^{-c}$, where $c = c(d, t, D)$. Let $R$ be the graph on $[r']$ with loops where $(i, j)$ is an edge of $R$ if and only if $Q_i, Q_j$ is complete to each other if $i \neq j$, and $(i, i)$ is an edge if the density in $Q_i$ is at least $1 - \alpha$. Round $A$ by the partition of $P$ and the graph $R$ to obtain another graph $A'$. That is, $A'$ has the same vertex set as $A$, and we delete the edges between $P_i$ and $P_j$ if $(i, j)$ is not an edge of $R$ and add all possible edges between $P_i$ and $P_j$ if $(i, j)$ is an edge of $R$. The resulting graph $A'$ is homogeneous between every pair of parts and at most an $\epsilon$-fraction of the pairs of vertices were added or deleted as edges from $A$ to obtain $A'$.

If $R \notin \mathcal{H}_r$, then every graph which has an induced homomorphism to $R$ is in $\mathcal{P}$ and hence $A' \in \mathcal{P}$. However, at most an $\epsilon$-fraction of the pairs were added or deleted from $A$ to obtain $A'$, and so $A$ is not $\epsilon$-far from satisfying $\mathcal{P}$, contradicting the assumption. Hence, $R \in \mathcal{H}_r$, and there is a graph $H \in \mathcal{H}$ on at most $s = \Psi_2(\mathcal{H}, r)$ vertices with $H \rightarrow_{\text{ind}} R$. Consider such an induced homomorphism $f : V(H) \rightarrow V(R)$, and let $a_i = |f^{-1}(i)|$. If among the sampled $v$ vertices there are at least $a_i$ vertices from $Q_i$ for each $i$, and these $a_i$ vertices form a clique if $(i, i)$ is a loop in $R$ and otherwise they form an independent set, then $H$ is an induced subgraph of the sampled set of vertices and hence the subgraph induced by the sampled set does not have property $\mathcal{P}$.

We first estimate the probability of the event that the sampled set of $v$ vertices contains $a_i$ vertices in $Q_i$ for each $i$. For a particular $i$, the probability that the sampled set of $v$ vertices contains fewer than $a = a_i$ vertices in $Q_i$ is 0 if $a = 0$ and is otherwise

$$\sum_{j=0}^{a-1} \binom{|Q_i|}{j} \binom{|V(A)| - |Q_i|}{v - j} \Big/ \binom{|V(A)|}{v} \quad < \quad (1 - \delta)^{v-a} \binom{v}{a} < e^{-\delta v/2} v^a < 1/(8s),$$

where we used that $a < s$ and $v$ can be chosen so that $v > 10(s/\delta)^2$. Taking the union bound and summing over all $i$, the probability that the sampled $v$ vertices does not contain $a_i$ vertices in $Q_i$ for some $i$ is at most $s \times 1/(8s) = 1/8$.

We now condition on the event that we have at least $a_i$ vertices chosen from $Q_i$. The probability that these $a_i$ vertices form a clique if $(i, i)$ is a loop and an independent set if $(i, i)$ is not a loop is at most $\binom{a_i}{2}\alpha$. Summing over all $i$, the probability that, for every $i$, the $a_i$ vertices in $Q_i$ form a clique if $(i, i)$ is a loop and an independent set if $(i, i)$ is not a loop, is at least

$$1 - \sum_i \binom{a_i}{2}\alpha \geq 1 - \binom{s}{2}\alpha \geq 7/8.$$

17

Hence, with probability at least $3/4$, the induced subgraph on the sampled set of $s$ vertices has the desired properties, which completes the proof. $\qquad\square$

Our goal for the rest of the subsection is to prove Theorem 5.3. We first prove a Ramsey-type lemma which states that semi-algebraic graphs contain large balanced complete or empty $h$-partite subgraphs.

**Lemma 5.4.** *For every $d, t$, and $D$, there is a constant $c = c(d, t, D)$ satisfying the following condition. For any positive integer $h$, any semi-algebraic graph $G = (P, E)$ in $d$-space with complexity $(t, D)$ has subsets $A_1, \ldots, A_h$ with $|A_1| = \cdots = |A_h| \geq h^{-c}|P|$ such that every pair $A_i, A_j$ with $i \neq j$ is complete or none of them are.*

*Proof.* For $h = 1$, the result is trivial by taking $A_1 = P$. Thus, we may assume $h \geq 2$. It is shown in [3] that every induced subgraph of $G$ on $h^C$ vertices contains a clique or an independent set of order $h$. Applying Theorem 1.3 with $\epsilon = \frac{1}{2h^C}$, we obtain an equitable partition with $\epsilon^{-O(1)}$ parts such that all but at most a $\frac{1}{2h^C}$-fraction of the pairs of parts are homogeneous. Applying Turán's theorem to the auxiliary graph with a vertex for each part and an edge between each homogeneous pair, we obtain $h^C$ parts that are pairwise homogeneous. Picking one vertex from each of these parts, we can obtain an induced subgraph with $h$ vertices which is complete or empty. The parts these vertices come from (after possibly deleting a vertex from some parts to guarantee that they have the same size) have the desired properties. $\qquad\square$

We next prove Theorem 5.3, a strengthening of our quantitative semi-algebraic regularity lemma, via three applications of Theorem 1.3.

*Proof of Theorem 5.3.* We will apply Theorem 1.3 three times. We first apply Theorem 1.3 to obtain a partition $P = P_1 \cup \cdots \cup P_K$ with $K = \epsilon^{-O(1)}$ with the implied constant depending on $d, t$, and $D$, such that all but an $\epsilon$-fraction of the pairs $P_i, P_j$ are homogeneous. We apply Theorem 1.3 again (or rather its proof) to get a refinement with approximation parameter $\epsilon' = 1/K^4$, so that all but an $\epsilon'$-fraction of the pairs of parts are homogeneous. Thus, with a positive probability, a random choice of parts $W_1, \ldots, W_K$ of this refinement with $W_i \subset P_i$ has the property that $W_i, W_j$ is homogeneous for all $i \neq j$. From Lemma 5.4, applied to the subgraph induced by $W_i$ for each $i$ with $h = 2/\alpha$, we obtain subsets $Q_i \subset W_i$ such that each $Q_i$ is a complete or empty balanced $h$-partite graph, so that the density in $Q_i$ is at most $\alpha$ or at least $1 - \alpha$. This completes the proof. $\square$

## 5.3 Testing hereditary properties of hypergraphs

We next state and prove the hereditary property testing result for semi-algebraic hypergraphs.

Let $R$ be a $k$-uniform hypergraph with vertex set $[r']$, and $B$ be a *blow-up* of $R$ with vertex sets $V_1, \ldots, V_{r'}$. That is, $B$ is a $k$-uniform hypergraph on $V_1 \cup \ldots \cup V_{r'}$, where $(v_1, \ldots, v_k) \in (V_{i_1}, \ldots, V_{i_k})$ is an edge if and only if $(i_1, \ldots, i_k)$ form an edge of $R$. An extension of $B$ (with respect to $V_1, \ldots, V_{r'}$) is any hypergraph on $V_1 \cup \ldots \cup V_{r'}$ which agrees with $B$ on the $k$-tuples with vertices in distinct $V_i$.

For a hypergraph $H$, we say that $R$ is *extendable $H$-free* if each blow-up of $R$ has an extension which contains no induced copy of $H$. For a family $\mathcal{H}$ of hypergraphs, we say that $R$ is *extendable $\mathcal{H}$-free* if each blow-up of $R$ has an extension which contains no induced $H \in \mathcal{H}$.

For a hypergraph property $\mathcal{P}$, we say that $R$ *strongly has property $\mathcal{P}$* if every blow-up of $R$ has an extension which has property $\mathcal{P}$. Otherwise, there are a smallest $s = s(\mathcal{P}, R)$ and vertex sets

$V_1, \ldots, V_{r'}$ with $r' \leq r$ and $s(\mathcal{P}, R) = |V_1| + \cdots + |V_{r'}|$ such that no extension of the blow-up $B$ of $R$ with vertex sets $V_1, \ldots, V_{r'}$ has property $\mathcal{P}$.

Define $\Psi_3(\mathcal{P}, r)$ to be the maximum of $s(\mathcal{P}, R)$ over all $R$ with at most $r$ vertices which do not strongly have property $\mathcal{P}$.

Our next theorem is about hereditary property testing for semi-algebraic hypergraphs. It implies that, if $\Psi_3(\mathcal{H}, r)$ is at most polynomial in $r$, then $\mathcal{P}$ can be easily tested within the semi-algebraic hypergraphs of constant description complexity, i.e., there is an $\epsilon$-tester with vertex query complexity $\epsilon^{-O(1)}$. A simple example in which $\Psi_3(\mathcal{H}, r)$ is constant is the case that the property $\mathcal{P}$ is the family of hypergraphs which are induced $H$-free for some fixed hypergraph $H$.

**Theorem 5.5.** *Let $\mathcal{P}$ be a hereditary property of hypergraphs. Within the family $\mathcal{A}$ of semi-algebraic $k$-uniform hypergraphs in $d$-space with description complexity $(t, D)$, the property $\mathcal{P}$ can be $\epsilon$-tested with vertex query complexity at most $r^C \Psi_3(\mathcal{H}, r)^2$, where $r = (1/\epsilon)^C$ with $C = C(d, t, D)$.*

We will need the following polynomial strong regularity lemma for semi-algebraic hypergraphs.

**Lemma 5.6.** *For any $0 < \alpha, \epsilon < 1/2$, any semi-algebraic $k$-uniform hypergraph $H = (P, E)$ in $d$-space with complexity $(t, D)$ has an equitable vertex partition $P = P_1 \cup \cdots \cup P_{r'}$ such that all but an $\epsilon$-fraction of the $k$-tuples of distinct parts are homogeneous. Furthermore, there are subsets $Q_i \subset P_i$ for each $i$ such that every $k$-tuple of distinct parts is homogeneous and $|Q_i| \geq \delta|P|$ with $\delta = \epsilon^{-c}$, where $c = c(k, d, t, D)$.*

*Proof.* The proof follows the graph case, as in Theorem 5.3, and involves two applications of Theorem 1.3. First, we apply Theorem 1.3 to obtain a partition $P = P_1 \cup \cdots \cup P_K$ with $K = \epsilon^{-O(1)}$, where the implied constant depends on $k, d, t, D$, such that all but an $\epsilon$-fraction of the $k$-tuples of parts are homogeneous. We apply Theorem 1.3 (or rather its proof) again to get a refinement with approximation parameter $\epsilon' = 1/K^{2k}$, so that all but an $\epsilon'$-fraction of the $k$-tuples of parts are homogeneous. Thus, with positive probability, a random choice of parts $Q_1, \ldots, Q_K$ of this refinement with $Q_i \subset P_i$ has the property that each $k$-tuple $Q_{i_1}, \ldots, Q_{i_k}$ of distinct parts is homogeneous. This completes the proof. $\square$

*Proof of Theorem 5.5.* Let $s = \Psi_3(\mathcal{P}, r)$ and $v = (rs)^C$ for an appropriate constant $C = C(k, d, t, D)$. Consider the tester which samples $v$ vertices from a graph $A = (P, E) \in \mathcal{A}$. It accepts if the induced subgraph on these $v$ vertices has property $\mathcal{P}$ and rejects otherwise. It suffices to show that if $A \in \mathcal{A}$ is $\epsilon$-far from satisfying $\mathcal{P}$, then with probability at least $2/3$, the tester will reject.

Consider an equitable partition $P = P_1 \cup \ldots \cup P_{r'}$ of the vertex set of $A$ guaranteed by Lemma 5.6 so that all but at most an $\epsilon$-fraction of the pairs of parts are homogeneous, and subsets $Q_i \subset P_i$ such that every $k$-tuple of distinct $Q_i$ is homogeneous, and $|Q_i| \geq \delta|P|$ for each $i$ with $\delta = \epsilon^{-c}$, where $c = c(k, d, t, D)$. Let $R$ be the $k$-uniform hypergraph on $[r']$, where a $k$-tuple $(i_1, \ldots, i_k)$ of distinct vertices forms an edge if and only if $Q_{i_1}, \ldots, Q_{i_k}$ is complete. If $R$ strongly has property $\mathcal{P}$, then $A$ is $\epsilon$-close to a hypergraph which has property $\mathcal{P}$, and hence the algorithm accepts in this case. Thus, we may assume that $R$ does not strongly have property $\mathcal{P}$ so that there are sets $V_1, \ldots, V_{r'}$ with $|V_1| + \cdots + |V_{r'}| = s(\mathcal{P}, R) \leq \Psi_3(\mathcal{P}, r) = s$ such that every extension of the blow-up of $R$ with parts $V_1, \ldots, V_{r'}$ does not have property $\mathcal{P}$. Thus, if among the $v$ sampled vertices, we get for every $i$ at least $|V_i|$ vertices in $Q_i$, then the subgraph induced by the sampled vertices does not have property $\mathcal{P}$ and the algorithm rejects. Therefore, it suffices to show that with probability at least $2/3$, we get at least $|V_i|$ vertices in each $Q_i$. However, this is the computation we already

did in the proof of Theorem 5.2. It is sufficient, for example, to assume that $v > 10(s/\delta)^2$, and we can take $v$ to satisfy this condition. $\qquad\square$

# References

[1] N. Alon, *Testing subgraphs in large graphs*, Random Structures Algorithms, **21** (2002), pp. 359–370.

[2] N. Alon and J. Fox, *Easily testable graph properties*, Combin. Probab. Comput., to appear.

[3] N. Alon, J. Pach, R. Pinchasi, R. Radoičić, and M. Sharir, *Crossing patterns of semi-algebraic sets*, J. Combin. Theory Ser. A, **111** (2005), pp. 310–326.

[4] N. Alon and A. Shapira, *A characterization of easily testable induced subgraphs*, Combin. Probab. Comput., **15** (2006), pp. 791–805.

[5] N. Alon and A. Shapira, *Every monotone graph property is testable*, SIAM J. Comput., **38** (2008), pp. 505–522.

[6] 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.

[7] T. Austin and T. Tao, *Testability and repair of hereditary hypergraph properties*, Random Structures Algorithms, **36** (2010), pp. 373–463.

[8] I. Bárány and J. Pach, *Homogeneous selections from hyperplanes*, J. Combinat. Theory Ser. B, **104** (2014), pp. 81–87.

[9] I. Bárány and P. Valtr, *A positive fraction Erdős-Szekeres theorem*, Discrete Comput. Geom., **19** (1998), pp. 335–342.

[10] S. Basu, R. Pollack, and M. F. Roy, *Algorithms in Real Algebraic Geometry*, 2nd Edition, Algorithms and Computation in Mathematics **10**, Springer-Verlag, Berlin, 2006.

[11] B. Bukh and A. Hubard, *Space crossing numbers*, Combin. Probab. Comput., **21** (2012), pp. 358–373.

[12] B. Chazelle, H. Edelsbrunner, L. Guibas, and M. Sharir, *A singly exponential stratification scheme for real semi-algebraic varieties and its applications*, Theor. Comput. Sci., **84** (1991), pp. 77–105.

[13] B. Chazelle and J. Friedman, *A deterministic view of random sampling and its use in geometry*, Combinatorica, **10** (1990), pp. 229–249.

[14] K. Clarkson, *A randomized algorithm for closest-point queries*, SIAM J. Comput., **17** (1988), pp. 830–847.

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

[16] D. Conlon, J. Fox, J. Pach, B. Sudakov, and A. Suk, *Ramsey-type results for semi-algebraic relations*, Trans. Amer. Math. Soc., **366** (2014), pp. 5043–5065.

[17] P. Erdős, *On extremal problems of graphs and generalized graphs*, Israel J. Math., **2** (1965), pp. 183–190.

[18] P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, Compos. Math., **2** (1935), pp. 463–470.

[19] J. Fox, M. Gromov, V. Lafforgue, A. Naor, and J. Pach, *Overlap properties of geometric expanders*, Reine Angew. Math. (Crelle's Journal), **671** (2012), pp. 49–83.

[20] J. Fox, J. Pach, and A. Suk, *Density and regularity theorems for semi-algebraic hypergraphs*, Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, 1517-1530, SIAM, San Diego, California, 2015.

[21] J. Fox and L. M. Lovász, *A tight lower bound for Szemerédi's regularity lemma*, preprint.

[22] O. Goldreich, S. Goldwasser, and D. Ron, *Property testing and its applications to learning and approximation*, J. ACM, **45** (1998), pp. 653–750.

[23] J. E. Goodman and R. Pollack, *Allowable sequences and order-types in discrete and computational geometry*, In J. Pach editor, New Trends in Discrete and Computational Geometry, **10** (1993), Springer, Berlin etc., pp. 103–134.

[24] J. E. Goodman and R. Pollack, *The complexity of point configurations*, Discrete Appl. Math., **31** (1991), pp. 167–180.

[25] W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math., **166** (2007), pp. 897–946.

[26] W. T. Gowers, *Lower bounds of tower type for Szemerédi's uniformity lemma*, Geom. Funct. Anal., **7** (1997), pp. 322–337.

[27] W. T. Gowers, *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combin. Probab. Comput., **15** (2006), pp. 143–184.

[28] A. Hubard, L. Montejano, E. Mora, and A. Suk, *Order types of convex bodies*, Order, **28** (2011), pp. 121–130.

[29] G. Kalai, *Intersection patterns of convex sets*, Israel J. Math., **48** (1984), pp. 161–174.

[30] R. Karasev, *A simpler proof of the Boros-Füredi-Bárány-Pach-Gromov theorem*, Discrete Comput. Geom., **47** (2012), pp. 492–495.

[31] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*. In D. Miklos et al. editors, Combinatorica, Paul Erdős Is Eighty, **2** (1996), pp. 295–352.

[32] T. Kővári, V. T. Sós, and P. Turán, *On a problem of K. Zarankiewicz*, Colloquium Math., **3** (1954), pp. 50–57.

[33] J. Kynčl, P. Paták, Z. Safernová, M. Tancer, Bounds for Pach's selection theorem and for the minimum solid angle in a simplex, arXiv:1403.8147

[34] J. Matoušek, *Lectures on Discrete Geometry*, Springer-Verlag New York, Inc., 2002.

[35] G. Moshkovitz and A. Shapira, *A short proof of Gowers' lower bound for the regularity lemma*, preprint.

[36] B. Nagle, V. Rödl and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures Algorithms, **28** (2006), pp. 113–179.

[37] J. Pach, *A Tverberg-type result on rainbow simplices*, Computational Geometry, **10** (1998), pp. 71–76.

[38] J. Pach and P. Agarwal, *Combinatorial Geometry*, New York, Wiley, 1995.

[39] V. Rödl and M. Schacht, *Property testing in hypergraphs and the removal lemma*, Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC 2007), ACM, New York, 2007, pp. 488–495.

[40] V. Rödl and M. Schacht, *Generalizations of the removal lemma*, Combinatorica, **29** (2009), pp. 467–501.

[41] R. Rubinfeld and M. Sudan, *Robust characterization of polynomials with applications to program testing*, SIAM J. on Computing, **25** (1996), pp. 252–271.

[42] E. Szemerédi, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay), **260**, CNRS, Paris, 1978, pp. 399–401.

[43] H. Tverberg, *A generalization of Radon's theorem*, J. Lond. Math. Soc., **41** (1966), pp. 123–128.

[44] R. T. Živaljević and S. T. Vrećica, *The colored Tverberg's problem and complexes of injective functions*, J. Combin. Theory Ser. A, **61** (1992), pp. 309–318.