



RESEARCH ARTICLE

# Geometrical sets with forbidden configurations

Davi Castro-Silva 

CWI & QuSoft, Science Park 123, 1098 XG Amsterdam, The Netherlands; E-mail: [davisilva15@gmail.com](mailto:davisilva15@gmail.com).

Received: 18 August 2021; Revised: 7 December 2022; Accepted: 24 April 2023

## Abstract

Given finite configurations  $P_1, \dots, P_n \subset \mathbb{R}^d$ , let us denote by  $\mathbf{m}_{\mathbb{R}^d}(P_1, \dots, P_n)$  the maximum density a set  $A \subseteq \mathbb{R}^d$  can have without containing congruent copies of any  $P_i$ . We will initiate the study of this geometrical parameter, called the *independence density* of the considered configurations, and give several results we believe are interesting. For instance we show that, under suitable size and nondegeneracy conditions,  $\mathbf{m}_{\mathbb{R}^d}(t_1 P_1, t_2 P_2, \dots, t_n P_n)$  progressively ‘untangles’ and tends to  $\prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(P_i)$  as the ratios  $t_{i+1}/t_i$  between consecutive dilation parameters grow large; this shows an exponential decay on the density when forbidding multiple dilates of a given configuration, and gives a common generalization of theorems by Bourgain and by Bukh in geometric Ramsey theory. We also consider the analogous parameter  $\mathbf{m}_{\mathbb{S}^d}(P_1, \dots, P_n)$  in the more complicated framework of sets on the unit sphere  $\mathbb{S}^d$ , obtaining the corresponding results in this setting.

## 1. Introduction

The general problem we consider in this paper can be phrased by the following question: how large can a set be if it does not contain a given geometrical configuration?

The simplest and most well-studied instance of this problem concerns forbidden configurations of only two points on  $\mathbb{R}^d$ , which are then characterized by their distance; since there clearly exist unbounded sets on  $\mathbb{R}^d$  which do not span a given distance, the appropriate notion of ‘largeness’ must take into account their density rather than their cardinality or measure. Define the *upper density*  $\bar{d}(A)$  of a measurable set  $A \subseteq \mathbb{R}^d$  by

$$\bar{d}(A) = \limsup_{T \rightarrow \infty} \frac{\text{vol}(A \cap [-T, T]^d)}{\text{vol}([-T, T]^d)},$$

where  $\text{vol}$  denotes the Lebesgue measure. Our general problem in this case becomes: what is the maximum upper density that a subset of  $\mathbb{R}^d$  can have if it does not contain pairs of points at distance 1?<sup>1</sup>

This extremal density is commonly denoted  $m_1(\mathbb{R}^d)$ , and it is associated to the measurable chromatic number<sup>2</sup>  $\chi_m(\mathbb{R}^d)$  of the Euclidean space by the simple inequality  $\chi_m(\mathbb{R}^d) \geq 1/m_1(\mathbb{R}^d)$ . Indeed, if no colour class contains pairs of points at unit distance, then each of them has upper density at most  $m_1(\mathbb{R}^d)$ , and it takes at least  $1/m_1(\mathbb{R}^d)$  such classes to cover the whole space. The parameter  $m_1(\mathbb{R}^d)$  is many times studied in the context of providing lower bounds for the measurable chromatic number.

Despite significant research on the subject, there is still no dimension  $d \geq 2$  for which the value of  $m_1(\mathbb{R}^d)$  is known. As far back as 1982, Erdős [9] conjectured that  $m_1(\mathbb{R}^2) < 1/4$ , implying that any

<sup>1</sup>Note that this problem is dilation invariant, so there is no loss of generality in assuming the forbidden distance to be 1.

<sup>2</sup>The measurable chromatic number of  $\mathbb{R}^d$  is the minimum number of measurable sets needed to partition  $\mathbb{R}^d$  so that no two points belonging to the same part are at distance 1 from each other.

measurable planar set covering one fourth of the Euclidean plane contains pairs of points at unit distance; this conjecture is still open. A celebrated theorem of Frankl and Wilson [11] implies that  $m_1(\mathbb{R}^d)$  decays exponentially with the dimension and obtains the asymptotic upper bound  $m_1(\mathbb{R}^d) \leq (1.2 + o(1))^{-d}$ . We refer the reader to Bachoc, Passuello and Thiery [1] and to DeCorte, Oliveira and Vallentin [6] for the best known bounds on  $m_1(\mathbb{R}^d)$  and  $\chi_m(\mathbb{R}^d)$ .

The situation becomes even more complex and interesting when one forbids multiple distances  $r_1, \dots, r_n > 0$ ; let us denote by  $\mathbf{m}_{\mathbb{R}^d}(r_1, \dots, r_n)$  the maximum upper density of a set in  $\mathbb{R}^d$  avoiding all of these distances. This parameter was first studied by Székely [22, 23] in connection with the chromatic number of geometric graphs, and it depends not only on the dimension of the space and number of forbidden distances but also on how these distances relate to each other.

In his first paper, Székely pondered on the connection between the structure of a set of forbidden distances and the maximum density of a set in Euclidean space which avoids them all, and conjectured that  $\mathbf{m}_{\mathbb{R}^2}((r_j)_{j \geq 1}) = 0$  whenever the sequence  $(r_j)_{j \geq 1}$  of forbidden distances is unbounded. His conjecture was proven by Furstenberg, Katznelson and Weiss [12] using methods from ergodic theory, who obtained the following result:

**Theorem 1.** *If  $A \subseteq \mathbb{R}^2$  has positive upper density, then there is some number  $t_0$  such that for any  $t \geq t_0$ , one can find a pair of points  $x, y \in A$  with  $\|x - y\| = t$ .*

Using Fourier analytic methods, Bourgain [2] was then able to generalize this theorem from two-point configurations on  $\mathbb{R}^2$  to  $d$ -point configurations in general position on  $\mathbb{R}^d$ , for any  $d \geq 2$ . For convenience, we shall say that a configuration  $P \subset \mathbb{R}^d$  is *admissible* if it has at most  $d$  points and spans a  $(|P| - 1)$ -dimensional affine hyperplane. Bourgain showed the following:

**Theorem 2.** *Suppose  $P \subset \mathbb{R}^d$  is admissible. If  $A \subseteq \mathbb{R}^d$  has positive upper density, then there is some number  $t_0 > 0$  such that  $A$  contains a congruent copy of  $t \cdot P$  for all  $t \geq t_0$ .*

This result motivates the introduction of the *independence density* of a given family of configurations  $P_1, P_2, \dots, P_n \subset \mathbb{R}^d$ , denoted  $\mathbf{m}_{\mathbb{R}^d}(P_1, P_2, \dots, P_n)$ , as the maximum upper density of a set in  $\mathbb{R}^d$  which does not contain a congruent copy of any of these configurations. This parameter generalizes our earlier notion of extremal density  $\mathbf{m}_{\mathbb{R}^d}(r_1, \dots, r_n)$  from two-point to higher-order configurations, and can be seen as the natural analogue of the independence number<sup>3</sup> for the (infinite) geometrical hypergraph on  $\mathbb{R}^d$  whose edges are all isometric copies of  $P_j$ ,  $1 \leq j \leq n$ .

With the notation now introduced, Bourgain's Theorem can be restated as the assertion that  $\mathbf{m}_{\mathbb{R}^d}((t_j P)_{j \geq 1}) = 0$  for all admissible  $P \subset \mathbb{R}^d$  and all unbounded positive sequences  $(t_j)_{j \geq 1}$ ; his proof, in fact, implies the stronger result that

$$\mathbf{m}_{\mathbb{R}^d}(t_1 P, t_2 P, \dots, t_n P) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

whenever the dilation parameters  $t_j$  grow without bound. Seen in this light, his results might inspire several further natural questions; for instance:

- (Q1) What is the rate of decay of  $\mathbf{m}_{\mathbb{R}^d}(t_1 P, t_2 P, \dots, t_n P)$  with  $n$  as the ratios  $t_{j+1}/t_j$  between consecutive scales get large?
- (Q2) What possible values can be taken by the independence density  $\mathbf{m}_{\mathbb{R}^d}(t_1 P, t_2 P, \dots, t_n P)$  of  $n$  distinct dilates of a given configuration  $P$ ?
- (Q3) Are there analogous results which are valid for other (non-Euclidean) spaces?

The goal of the present paper is to initiate the study of the independence density function  $\mathbf{m}_{\mathbb{R}^d}$  and related geometrical parameters, and the investigation of these three problems will serve as the driving force behind our analysis.

<sup>3</sup>Given some finite hypergraph  $H$ , its independence number is the maximum size of a subset of vertices which does not entirely contain any edge. Its independence density can then be defined as the independence number divided by the total number of vertices.

### 1.1. Outline of the paper

In Section 2, we will formally define the independence density of a family of configurations, both in the entire space  $\mathbb{R}^d$  and when restricted to bounded cubes in  $\mathbb{R}^d$ , and start our study of this geometrical parameter. The methods we use are a mix of Fourier analysis, functional analysis and combinatorics. The Fourier-analytic part is based mainly on Bourgain's arguments from [2], and the combinatorial part is based on Bukh's arguments from [3] (where he considered similar problems to ours but concerning forbidden distances). We do not assume that the reader is familiar with either of these papers; instead, we give a presentation of the relevant parts of their reasoning that will be important to us.

The main tools to be used in this section will be a Counting Lemma (Lemma 4) and a Supersaturation Theorem (Theorem 3), both of which are conceptually similar to results of the same name in graph and hypergraph theory (see [18, 4, 10]). Intuitively, the Counting Lemma says that the count of admissible configurations inside a given set does not significantly change if we blur the set a little; this will be proven by Fourier-analytic methods. The Supersaturation Theorem states that any bounded set  $A \subseteq [-R, R]^d$ , which is just slightly denser than the independence density of an admissible configuration  $P$ , must necessarily contain a positive proportion of all congruent copies of  $P$  lying in  $[-R, R]^d$ ; this is proven by functional-analytic methods, via a compactness and weak\* continuity argument.

We will then use these tools to obtain several results on the independence density parameter, and in particular, answer questions (Q1) and (Q2) in the case where the considered configuration  $P$  is admissible. Regarding question (Q1), we show that  $\mathbf{m}_{\mathbb{R}^d}(t_1P, t_2P, \dots, t_nP)$  tends to  $\mathbf{m}_{\mathbb{R}^d}(P)^n$  as the ratios  $t_{j+1}/t_j$  get large; this generalizes a theorem of Bukh from two-point configurations to  $k$ -point configurations with  $k \leq d$  and easily implies Bourgain's Theorem discussed in the Introduction. As for question (Q2), we show that, by forbidding  $n$  distinct dilates of such a configuration  $P$ , we can obtain as independence density any real number strictly<sup>4</sup> between  $\mathbf{m}_{\mathbb{R}^d}(P)^n$  and  $\mathbf{m}_{\mathbb{R}^d}(P)$ , but none smaller than  $\mathbf{m}_{\mathbb{R}^d}(P)^n$  or larger than  $\mathbf{m}_{\mathbb{R}^d}(P)$ . We also prove:

- The general lower bound  $\mathbf{m}_{\mathbb{R}^d}(P_1, P_2, \dots, P_n) \geq \prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(P_i)$ , which holds for all configurations  $P_1, P_2, \dots, P_n \subset \mathbb{R}^d$ ;
- Continuity of the independence density function  $\mathbf{m}_{\mathbb{R}^d}$  on the set of admissible configurations; and
- Existence of extremizer measurable sets (i.e., having maximal density) which avoid admissible configurations.

In Section 3, we will consider these same questions but related to the more complicated setting of sets on the unit sphere  $\mathbb{S}^d$ . We will also present (and prove) a spherical analogue of Bourgain's Theorem; this is in line with our question (Q3), as the sphere is the most well-studied non-Euclidean space.

Many of the arguments from the Euclidean setting will be used again in the spherical setting (in particular, the reliance on our two main combinatorial tools), but there are also some complications we need to solve that are intrinsic to the sphere. One of them is that harmonic analysis is (for our purposes) much more complicated on  $\mathbb{S}^d$  than it is on  $\mathbb{R}^d$ , which makes our proof of the spherical Counting Lemma correspondingly harder and more technical than its Euclidean counterpart. Moreover, due to the lack of dilation invariance in the spherical setting, we will only be able to make a modest progress towards answering its analogue of question (Q2) (and the answer to question (Q1) will be somewhat more intricate). The other results proven in the Euclidean space setting will continue to hold in the same form for sets on the sphere.

Finally, in Section 4, we discuss some related results in the literature and suggest several intriguing open problems in line with the results presented here.

### 1.2. Some remarks on notation

The same denomination will be used for both a set and its indicator function; for instance, if we are given  $A \subseteq \mathbb{R}^d$ , then  $A(x) = 1$  if  $x \in A$  and  $A(x) = 0$  otherwise. The group of permutations of  $\{1, \dots, k\}$

<sup>4</sup>Whether these boundary values can be attained is not yet clear.

is denoted by  $\mathfrak{S}_k$ . Given a group  $G$  acting on some space  $X$  and an element  $x$  of this space, we write  $\text{Stab}^G(x) := \{g \in G : g.x = x\}$  for the stabilizer subgroup of  $x$ .

The averaging notation  $\mathbb{E}_{x \in X}$  is used to denote the expectation when the variable  $x$  is distributed uniformly over the set  $X$ . When  $X$  is (a subset of) a compact group  $G$ , this measure is (the restriction of) the normalized Haar measure on  $G$ , which is the unique Borel probability measure on  $G$  which is invariant by both left- and right-actions of this group. Similarly, we write  $\mathbb{P}_{x \in X}$  to denote the probability under this same distribution.

## 2. Configurations in Euclidean space

Throughout this section, we shall fix an integer  $d \geq 2$  and work on the  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ , equipped with its usual inner product  $x \cdot y$  and associated Euclidean norm  $\|x\|$ . We denote by  $\text{vol}$  the Lebesgue measure on  $\mathbb{R}^d$  and by  $\mu$  the normalized Haar measure on the orthogonal group  $O(\mathbb{R}^d) = \{O \in \mathbb{R}^{d \times d} : O^t O = I\}$ .

Given  $x \in \mathbb{R}^d$  and  $R > 0$ , we denote by  $Q(x, R)$  the axis-parallel open cube of side length  $R$  centered at  $x$ . We write  $d_{Q(x,R)}(A) := \text{vol}(A \cap Q(x, R))/R^d$  for the density of  $A \subseteq \mathbb{R}^d$  inside the cube  $Q(x, R)$ . The upper density of a measurable set  $A \subseteq \mathbb{R}^d$  can then be written as  $\bar{d}(A) = \limsup_{R \rightarrow \infty} d_{Q(0,R)}(A)$ ; if the limit exists, we shall instead denote it by  $d(A)$ .

A *configuration*  $P$  is just a finite subset of  $\mathbb{R}^d$ , and we define its *diameter*  $\text{diam } P$  as the largest distance between two of its points. Recall that a configuration  $P \subset \mathbb{R}^d$  on  $k$  points is said to be *admissible* if  $k \leq d$  and if  $P$  is nondegenerate (that is, if it spans a  $(k - 1)$ -dimensional affine hyperplane). The space of  $k$ -point configurations can be given a metric induced from the Euclidean norm as follows: if  $P = \{v_1, \dots, v_k\}$  and  $Q = \{u_1, \dots, u_k\}$ , the distance between  $P$  and  $Q$  is

$$\|P - Q\|_\infty := \min_{\sigma \in \mathfrak{S}_k} \max_{1 \leq i \leq k} \|v_i - u_{\sigma(i)}\|,$$

where the minimum is taken over all permutations  $\sigma$  of  $\{1, \dots, k\}$ . It is easy to see that, under the topology induced by this metric, the set of admissible configurations is an open set and that it is dense inside the family of all subsets of  $\mathbb{R}^d$  with at most  $d$  elements.

We say that two configurations  $P, Q \subset \mathbb{R}^d$  are *congruent*, and write  $P \simeq Q$ , if they can be made equal using only rigid transformations; that is,  $P \simeq Q$  if and only if there exist  $x \in \mathbb{R}^d$  and  $T \in O(\mathbb{R}^d)$  such that  $P = x + T \cdot Q$ . Given a configuration  $P \subset \mathbb{R}^d$ , we say that a set  $A \subseteq \mathbb{R}^d$  *avoids*  $P$  if there is no subset of  $A$  which is congruent to  $P$ .

We can now formally define our main object of study in this section, the *independence density* of a configuration or family of configurations. There are, in fact, two closely related versions of this parameter we will need, depending on whether we are considering bounded or unbounded configuration-avoiding sets. Given  $n \geq 1$  configuration  $P_1, \dots, P_n \subset \mathbb{R}^d$ , we then define the quantities

$$\begin{aligned} \mathbf{m}_{\mathbb{R}^d}(P_1, \dots, P_n) &:= \sup \{ \bar{d}(A) : A \subset \mathbb{R}^d \text{ avoids } P_i, 1 \leq i \leq n \} \text{ and} \\ \mathbf{m}_{Q(0,R)}(P_1, \dots, P_n) &:= \sup \{ d_{Q(0,R)}(A) : A \subset Q(0, R) \text{ avoids } P_i, 1 \leq i \leq n \}. \end{aligned}$$

These parameters are analogous to the notion of independence number of a hypergraph: if we consider the hypergraph on vertex set  $\mathbb{R}^d$  (resp.  $Q(0, R)$ ) whose edges are all isometric copies of  $P_j$ ,  $1 \leq j \leq n$ , then  $\mathbf{m}_{\mathbb{R}^d}(P_1, \dots, P_n)$  (resp.  $\mathbf{m}_{Q(0,R)}(P_1, \dots, P_n)$ ) can be thought of as the density of a largest independent set in this hypergraph.

**Remark.** For the sake of clarity and notational convenience, whenever possible the results we give about independence density will be stated and proved in the case of only one forbidden configuration. It can be easily verified that these results also hold in the case of several (but finitely many) forbidden configurations, with essentially unchanged proofs. Whenever we need this greater generality we will mention how the corresponding statement would be in the case of several configurations.

We start our investigations by proving a simple lemma which relates the two versions of independence density just defined.

**Lemma 1.** For all configurations  $P \subset \mathbb{R}^d$  and all  $R > 0$ , we have

$$\frac{\mathbf{m}_{Q(0,R)}(P)}{\left(1 + \frac{\text{diam } P}{R}\right)^d} \leq \mathbf{m}_{\mathbb{R}^d}(P) \leq \mathbf{m}_{Q(0,R)}(P).$$

*Proof.* For the first inequality, suppose  $A \subseteq Q(0, R)$  is a set avoiding  $P$  and consider the periodic set  $A' := A + (R + \text{diam } P)\mathbb{Z}^d$ . This set also avoids  $P$ , and it has density

$$d(A') = \frac{\text{vol}(A)}{(R + \text{diam } P)^d} = \frac{d_{Q(0,R)}(A)}{\left(1 + \frac{\text{diam } P}{R}\right)^d}.$$

Since we can choose  $d_{Q(0,R)}(A)$  arbitrarily close to  $\mathbf{m}_{Q(0,R)}(P)$ , the leftmost inequality follows.

Now, let  $A \subseteq \mathbb{R}^d$  be any set avoiding  $P$  and note that  $A \cap Q(x, R)$  also avoids  $P$  for every  $x \in \mathbb{R}^d$ . By fixing  $\varepsilon > 0$  and then averaging over all  $x$  inside a large enough cube  $Q(0, R')$  (depending on  $A$ ,  $\text{diam } P$  and  $\varepsilon$ ), we conclude there is  $x \in \mathbb{R}^d$  for which  $\text{vol}(A \cap Q(x, R)) > (d(A) - \varepsilon)R^d$ . The rightmost inequality follows.  $\square$

As we are interested in the study of sets avoiding certain configurations, it is useful to also have a way of *counting* how many such configurations there are in a given set. For a given configuration  $P = \{v_1, v_2, \dots, v_k\} \subset \mathbb{R}^d$  and a measurable set  $A \subseteq \mathbb{R}^d$ , we define

$$I_P(A) := \int_{\mathbb{R}^d} \int_{O(\mathbb{R}^d)} A(x + Tv_1)A(x + Tv_2) \cdots A(x + Tv_k) d\mu(T) dx,$$

which represents how many (congruent) copies of  $P$  are contained in  $A$ . This quantity  $I_P(A)$  can, of course, be infinite if the set  $A$  is unbounded, but we will use it almost exclusively for bounded sets. We can similarly define its weighted version

$$I_P(f) := \int_{\mathbb{R}^d} \int_{O(\mathbb{R}^d)} f(x + Tv_1)f(x + Tv_2) \cdots f(x + Tv_k) d\mu(T) dx,$$

whenever  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is a measurable function for which this integral makes sense (say, for  $f \in L^k(\mathbb{R}^d)$ ). A large part of our analysis consists of getting a better understanding of the counting function  $I_P$ .

When a measurable set  $A \subseteq \mathbb{R}^d$  avoids some configuration  $P$ , it is clear from the definition that  $I_P(A) = 0$ ; however, it is also possible for  $I_P(A)$  to be zero even when  $A$  contains congruent copies of  $P$ . In intuitive terms, the condition  $I_P(A) = 0$  means only that  $A$  contains a negligible fraction of all possible copies of  $P$ . The next result shows that this distinction is essentially irrelevant for most purposes.

**Lemma 2** (Zero-measure removal). Suppose  $P \subset \mathbb{R}^d$  is a finite configuration and  $A \subseteq \mathbb{R}^d$  is measurable. If  $I_P(A) = 0$ , then we can remove a zero-measure subset of  $A$  in order to remove all copies of  $P$ .

*Proof.* By the Lebesgue Density Theorem, we have that

$$\lim_{\delta \rightarrow 0} \left| \frac{1}{\delta^d} \int_{Q(x,\delta)} A(y) dy - A(x) \right| = 0 \quad \text{for almost every } x \in \mathbb{R}^d.$$

Now, we remove from  $A$  all points  $x$  for which this identity does not hold, thus obtaining a subset  $B \subseteq A$

with  $\text{vol}(A \setminus B) = 0$  and

$$\lim_{\delta \rightarrow 0} \frac{1}{\delta^d} \int_{Q(x, \delta)} B(y) \, dy = 1 \quad \text{for all } x \in B.$$

We will show that no congruent copy of  $P$  remains on this restricted set  $B$ .

Suppose, for contradiction, that  $B$  contains a copy  $\{u_1, \dots, u_k\}$  of  $P$ . By assumption, there exists some  $\delta > 0$  such that

$$d_{Q(u_i, \delta)}(B) = \frac{1}{\delta^d} \int_{Q(u_i, \delta)} B(y) \, dy \geq 1 - \frac{1}{2^{d+1}k} \quad \text{for all } 1 \leq i \leq k; \tag{1}$$

fix such a value of  $\delta$ . Note that, if  $d_{Q(x, \delta)}(B) \geq 1 - 1/(2^{d+1}k)$  for some  $x \in \mathbb{R}^d$ , then for all  $y \in Q(x, \delta/2)$ , we have

$$\begin{aligned} d_{Q(y, \delta/2)}(B) &= 1 - \frac{\text{vol}(Q(y, \delta/2) \setminus B)}{(\delta/2)^d} \\ &\geq 1 - \frac{\text{vol}(Q(x, \delta) \setminus B)}{(\delta/2)^d} \\ &= 1 - \frac{\delta^d(1 - d_{Q(x, \delta)}(B))}{(\delta/2)^d} \\ &\geq 1 - \frac{1}{2k}. \end{aligned}$$

Our hypothesis (1) thus implies that  $d_{Q(y, \delta/2)}(B) \geq 1 - 1/2k$  whenever  $y \in Q(u_i, \delta/2)$  for some  $1 \leq i \leq k$ .

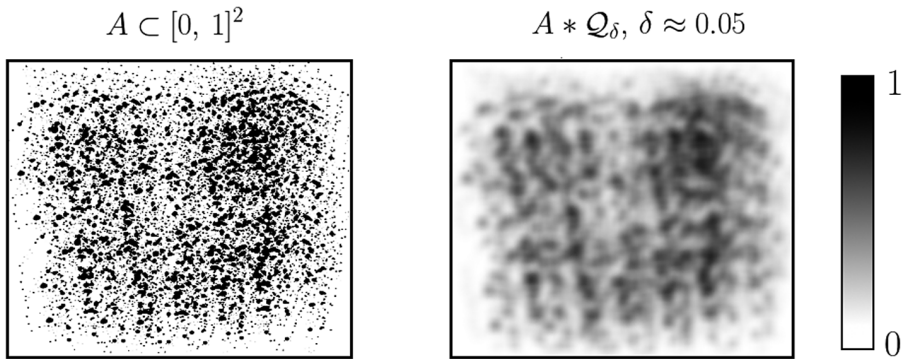
Let  $\ell := \max\{\|u_i\| : 1 \leq i \leq k\}$  be the largest length of a vector in our copy of  $P$ , and let us write  $\mathcal{B}(I, \delta/(4\ell)) := \{T \in O(\mathbb{R}^d) : \|T - I\| \leq \delta/(4\ell)\}$  for the ball of radius  $\delta/(4\ell)$  in spectral norm centered on the identity  $I$ . Note that, whenever  $T \in \mathcal{B}(I, \delta/(4\ell))$ , we have that  $Tu_i \in Q(u_i, \delta/2)$  for each  $1 \leq i \leq k$ . By the union bound, we then have

$$\begin{aligned} \int_{\mathbb{R}^d} \prod_{i=1}^k B(x + Tu_i) \, dx &\geq \int_{Q(0, \delta/2)} \prod_{i=1}^k B(x + Tu_i) \, dx \\ &= \left(\frac{\delta}{2}\right)^d \mathbb{P}_{x \in Q(0, \delta/2)}(x + Tu_i \in B \text{ for all } 1 \leq i \leq k) \\ &\geq \left(\frac{\delta}{2}\right)^d \left(1 - \sum_{i=1}^k \mathbb{P}_{x \in Q(0, \delta/2)}(x + Tu_i \notin B)\right) \\ &= \left(\frac{\delta}{2}\right)^d \left(1 - \sum_{i=1}^k (1 - d_{Q(Tu_i, \delta/2)}(B))\right) \\ &\geq \frac{1}{2} \left(\frac{\delta}{2}\right)^d. \end{aligned}$$

This immediately implies that

$$I_P(B) \geq \int_{\mathbb{R}^d} \int_{\mathcal{B}(I, \delta/(4\ell))} \prod_{i=1}^k B(x + Tu_i) \, d\mu(T) \, dx \geq \frac{\mu(\mathcal{B}(I, \delta/(4\ell)))}{2} \left(\frac{\delta}{2}\right)^d > 0,$$

contradicting our assumption that  $I_P(A) = 0$  and finishing the proof. □



**Figure 1.** An example of a planar set  $A$  on the unit square and the corresponding function  $A * Q_\delta$ , for some small  $\delta$ ; the shades of gray represent the value this function takes at each point.

**2.1. Fourier analysis on  $\mathbb{R}^d$  and the Counting Lemma**

We next show that the count of copies of an admissible configuration  $P$  inside a measurable set  $A$  does not significantly change if we ignore its fine details and ‘blur’ the set  $A$  a little. The philosophy is similar to the famous *regularity method* in graph theory, where a large graph can be replaced by a much smaller weighted ‘reduced graph’ (which is an averaged version of the original graph which ignores its fine details) without significantly changing the count of copies of any small subgraph.

The methods we will use are Fourier analytic in nature, drawing from Bourgain’s arguments presented in [2]. We define the Fourier transform on  $\mathbb{R}^d$  by

$$\widehat{f}(\xi) := \int_{\mathbb{R}^d} f(x)e^{-2\pi i x \cdot \xi} dx \quad \text{and} \quad \widehat{\sigma}(\xi) := \int_{\mathbb{R}^d} e^{-2\pi i x \cdot \xi} d\sigma(x)$$

for a (complex-valued) function  $f \in L^1(\mathbb{R}^d)$  and a finite Borel measure  $\sigma$  on  $\mathbb{R}^d$ . The convolution between two functions  $f, g \in L^1(\mathbb{R}^d)$  is defined by

$$f * g(x) := \int_{\mathbb{R}^d} f(y)g(x - y) dy.$$

We recall the basic identities  $\widehat{f * g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi)$  and

$$\int_{\mathbb{R}^d} f(x) d\sigma(x) = \int_{\mathbb{R}^d} \widehat{f}(\xi)\widehat{\sigma}(-\xi) d\xi,$$

as well as Parseval’s Identity  $\|f\|_2 = \|\widehat{f}\|_2$  for  $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ . For background in Fourier analysis, we refer the reader to the classic textbook of Stein and Weiss [20].

Denote  $Q_\delta(x) := \delta^{-d}Q(0, \delta)(x)$ . This way,  $f * Q_\delta(x) = \delta^{-d} \int_{Q(x, \delta)} f(y) dy$  is the average of a function  $f$  on the cube  $Q(x, \delta)$ . Specializing to the indicator function of a set  $A \subseteq \mathbb{R}^d$ , we obtain  $A * Q_\delta(x) = d_{Q(x, \delta)}(A)$ ; this represents a ‘blurring’ of the set  $A$  considered (see Figure 1). What we wish to obtain is then an upper bound on the difference  $|I_P(A) - I_P(A * Q_\delta)|$  which goes to zero as  $\delta$  goes to zero uniformly over all measurable sets  $A \subseteq Q(0, R)$  (for any fixed  $R > 0$ ).

Before delving into the details of our argument, let us present a simple telescoping sum argument which will be needed here and will be reused several times in this paper. Suppose we wish to bound



from above the expression

$$|I_P(f) - I_P(g)| = \left| \int_{\mathbb{R}^d} \int_{O(\mathbb{R}^d)} \left( \prod_{i=1}^k f(x + Tv_i) - \prod_{i=1}^k g(x + Tv_i) \right) d\mu(T) dx \right|$$

for some given functions  $f, g$  and some configuration  $P = \{v_1, \dots, v_k\}$ . Since we can rewrite the term inside the parenthesis above as the telescoping sum

$$\sum_{i=1}^k \left( \prod_{j=1}^{i-1} f(x + Tv_j) \right) (f(x + Tv_i) - g(x + Tv_i)) \left( \prod_{j=i+1}^k g(x + Tv_j) \right),$$

it follows from the triangle inequality that  $|I_P(f) - I_P(g)|$  is at most

$$\sum_{i=1}^k \left| \int_{\mathbb{R}^d} \int_{O(\mathbb{R}^d)} \prod_{j=1}^{i-1} f(x + Tv_j) (f(x + Tv_i) - g(x + Tv_i)) \prod_{j=i+1}^k g(x + Tv_j) d\mu(T) dx \right|.$$

To obtain some bound for  $|I_P(f) - I_P(g)|$ , it then suffices to obtain a similar bound for an expression of the form

$$\left| \int_{\mathbb{R}^d} \int_{O(\mathbb{R}^d)} h_1(x + Tu_1) \cdots h_{k-1}(x + Tu_{k-1}) (f(x + Tu_k) - g(x + Tu_k)) d\mu(T) dx \right|$$

whenever each  $h_i$  is either  $f$  or  $g$ , and whenever  $(u_1, \dots, u_k)$  is a permutation of the points of  $P$ .

We shall refer to an argument of this form (breaking a difference of products into a telescoping sum, using the triangle inequality and bounding each term of the resulting expression) as the *telescoping sum trick*. It is frequently used in modern graph and hypergraph theory when estimating the number of subgraphs inside a given large (hyper)graph  $G$  with the aid of edge-discrepancy measures such as the cut norm; such results are usually known as *counting lemmas* and are an essential part of the regularity method we have already mentioned (see the surveys [18, 4] for details).

In our arguments, we will also need some analytic facts and estimates, which we now provide. Given an  $m$ -dimensional subspace  $U \subseteq \mathbb{R}^d$ , we denote by  $\sigma_U^{(m-1)}$  the uniform probability measure on its unit sphere  $\mathbb{S}_U^{m-1} := \{x \in U : \|x\| = 1\}$ . This measure is closely related to the Haar measure  $\mu_U$  on the orthogonal group  $O(U)$ : if  $X \subseteq \mathbb{S}_U^{m-1}$  is a measurable set and  $x \in \mathbb{S}_U^{m-1}$  is any point, then

$$\sigma_U^{(m-1)}(X) = \mu_U(\{T \in O(U) : Tx \in X\})$$

(see, for instance, [5, Appendix A.5] for a simple proof of this fact). Given  $T \in O(\mathbb{R}^d)$ , we write  $TU := \{Tu : u \in U\}$  for the rotated subspace.

**Lemma 3.** *There are constants  $C_1, C_2 > 0$  (depending on the dimension  $d$ ) such that*

$$|1 - \widehat{Q}_\delta(\xi)| \leq C_1 \delta^2 \|\xi\|^2 \quad \text{for all } \delta > 0, \xi \in \mathbb{R}^d$$

and, if  $V$  is an  $m$ -dimensional subspace of  $\mathbb{R}^d$ ,

$$\int_{O(\mathbb{R}^d)} |\widehat{\sigma}_{TV}^{(m-1)}(\xi)|^2 d\mu(T) \leq C_2 \|\xi\|^{-(m-1)} \quad \text{for all } \xi \in \mathbb{R}^d \setminus \{0\}.$$



*Proof.* For the first inequality, note that

$$\widehat{Q}_\delta(\xi) = \frac{1}{\delta^d} \int_{Q(0,\delta)} e^{-2\pi i x \cdot \xi} dx = \prod_{j=1}^d \frac{1}{\delta} \int_{-\delta/2}^{\delta/2} e^{-2\pi i x_j \xi_j} dx_j = \prod_{j=1}^d \frac{\sin(\pi \delta \xi_j)}{\pi \delta \xi_j}$$

(where the  $j$ -th term in the product is 1 if  $\xi_j = 0$ ). It follows from the Taylor expansion of  $\sin(\cdot)$  that  $|x - \sin(x)| \leq C|x|^3$  for some  $C > 0$  and all  $x \in [-1, 1]$ . As the sine function is bounded, we conclude there is some constant  $C_1 > 0$  (depending on  $d$ ) for which

$$|1 - \widehat{Q}_\delta(\xi)| = \left| 1 - \prod_{j=1}^d \frac{\sin(\pi \delta \xi_j)}{\pi \delta \xi_j} \right| \leq C_1 \sum_{j=1}^d (\delta \xi_j)^2 = C_1 \delta^2 \|\xi\|^2$$

holds for all  $\delta > 0, \xi \in \mathbb{R}^d$ .

For the second inequality, we use the estimate

$$|\widehat{\sigma}_U^{(m-1)}(\xi)| \leq K \|\pi_U \xi\|^{-(m-1)/2},$$

where  $\pi_U \xi$  is the orthogonal projection of  $\xi$  onto  $U$  and  $K$  is an absolute constant. This estimate follows from

$$\widehat{\sigma}_U^{(m-1)}(\xi) = \int_{\mathbb{R}^d} e^{-2\pi i x \cdot \xi} d\sigma_U^{(m-1)}(x) = \int_U e^{-2\pi i x \cdot \pi_U \xi} d\sigma_U^{(m-1)}(x)$$

and the well-known asymptotic bound  $|\widehat{\sigma}_{\mathbb{R}^m}^{(m-1)}(\xi)| = O(\|\xi\|^{-(m-1)/2})$  for the unit sphere on  $\mathbb{R}^m$  (see, for instance, Chapter VIII, Section 3 in Stein's book [19]). For any  $\xi \in \mathbb{R}^d \setminus \{0\}$ , we then have that

$$\begin{aligned} \int_{O(\mathbb{R}^d)} |\widehat{\sigma}_{TV}^{(m-1)}(\xi)|^2 d\mu(T) &\leq \int_{O(\mathbb{R}^d)} K^2 \|\pi_{TV} \xi\|^{-(m-1)} d\mu(T) \\ &= K^2 \int_{O(\mathbb{R}^d)} \|\pi_V(T^{-1}\xi)\|^{-(m-1)} d\mu(T) \\ &= K^2 \|\xi\|^{-(m-1)} \int_{\mathbb{S}^{d-1}} \|\pi_{\mathbb{R}^m} y\|^{-(m-1)} d\sigma_{\mathbb{R}^d}^{(d-1)}(y), \end{aligned}$$

where we performed the change of variables  $y = T^{-1}\xi/\|\xi\|$ .

It now suffices to show that the last integral above is finite, which we will do by induction on  $d \geq m$ . In the base case where  $d = m$ , the integral is clearly equal to 1 since the projection operator  $\pi_{\mathbb{R}^m}$  is the identity. If  $d \geq m + 1$ , parameterize  $\mathbb{S}^{d-1}$  by

$$y = (z \sin \theta, \cos \theta) \quad \text{for } z \in \mathbb{S}^{d-2}, \theta \in [0, \pi];$$

denoting by  $\omega_{d-1}$  (resp.  $\omega_{d-2}$ ) the total Lebesgue measure of the unit sphere of  $\mathbb{R}^d$  (resp.  $\mathbb{R}^{d-1}$ ), this change of variables gives

$$\omega_{d-1} d\sigma_{\mathbb{R}^d}^{(d-1)}(y) = \omega_{d-2} (\sin \theta)^{d-2} d\sigma_{\mathbb{R}^{d-1}}^{(d-2)}(z) d\theta.$$

We then obtain

$$\begin{aligned} & \int_{\mathbb{S}^{d-1}} \|\pi_{\mathbb{R}^m} y\|^{-(m-1)} d\sigma_{\mathbb{R}^d}^{(d-1)}(y) \\ &= \frac{1}{\omega_{d-1}} \int_0^\pi \int_{\mathbb{S}^{d-2}} (\sin \theta \|\pi_{\mathbb{R}^m} z\|)^{-(m-1)} \omega_{d-2}(\sin \theta)^{d-2} d\sigma_{\mathbb{R}^{d-1}}^{(d-2)}(z) d\theta \\ &= \frac{\omega_{d-2}}{\omega_{d-1}} \left( \int_0^\pi (\sin \theta)^{d-m-1} d\theta \right) \int_{\mathbb{S}^{d-2}} \|\pi_{\mathbb{R}^m} z\|^{-(m-1)} d\sigma_{\mathbb{R}^{d-1}}^{(d-2)}(z) \\ &\leq \frac{\omega_{d-2}\pi}{\omega_{d-1}} \int_{\mathbb{S}^{d-2}} \|\pi_{\mathbb{R}^m} z\|^{-(m-1)} d\sigma_{\mathbb{R}^{d-1}}^{(d-2)}(z), \end{aligned}$$

and the desired bound follows by induction. □

We are now ready to formally state and prove our main technical tool in the Euclidean setting, which by analogy with methods from graph theory we shall call the Counting Lemma. We note that the main steps of its proof were already present in Bourgain’s paper [2].

**Lemma 4** (Counting Lemma). *For every admissible configuration  $P \subset \mathbb{R}^d$ , there exists a constant  $C_P > 0$  such that the following holds: for every  $R > 0$  and any measurable set  $A \subseteq Q(0, R)$ , we have that*

$$|I_P(A) - I_P(A * Q_\delta)| \leq C_P \delta^{1/4} R^d \quad \text{for all } \delta \in (0, 1].$$

Moreover, the same constant  $C_P$  can be made to hold uniformly over all configurations  $P'$  inside a neighborhood of  $P$ .

*Proof.* Let  $(v_1, \dots, v_k)$  be a fixed permutation of the points of  $P$ . We will work a bit more generally and show that a bound as in the statement of the lemma holds for

$$\left| \int_{\mathbb{R}^d} \int_{O(\mathbb{R}^d)} f_1(x + Tv_1) \cdots f_{k-1}(x + Tv_{k-1}) (f_k(x + Tv_k) - f_k * Q_\delta(x + Tv_k)) d\mu(T) dx \right|$$

whenever  $f_1, \dots, f_k : Q(0, R) \rightarrow [-1, 1]$  are measurable functions. By our telescoping sum trick, this immediately implies the result.

We first exploit the translation invariance of the problem in order to simplify the argument later on. Let  $U \subset \mathbb{R}^d$  denote the  $(k - 2)$ -dimensional affine hyperplane spanned by  $v_1, \dots, v_{k-1}$ , and let  $\pi_U v_k$  be the orthogonal projection of  $v_k$  onto  $U$  (so  $\pi_U v_k$  is the point in  $U$  which is closest to  $v_k$ ). By translating all points in  $P$  by  $-\pi_U v_k$ , we may assume that  $U$  contains the origin (being thus a subspace of  $\mathbb{R}^d$ ) and that  $v_k$  belongs to its orthogonal complement  $U^\perp$ . Note that  $v_k \neq 0$  since the points in  $P$  are affinely independent, and  $U^\perp$  has dimension  $d - k + 2 \geq 2$ ; these are the two properties we will need in the proof which require the assumption that  $P$  is admissible.

Let  $H := \text{Stab}^{O(\mathbb{R}^d)}(U)$  denote the subgroup of orthogonal transformations which act trivially on the subspace  $U$  and let  $\nu_H$  be the Haar measure on  $H$ . Let  $G := f_k - f_k * Q_\delta$  and, for a given  $T \in O(\mathbb{R}^d)$ , define the function  $F_T : Q(0, R) \rightarrow [-1, 1]$  by

$$F_T(x) = \prod_{i=1}^{k-1} f_i(x + Tv_i).$$

The integrand on the expression we wish to bound can then be written more succinctly as  $F_T(x)G(x +$

$Tv_k$ ). By symmetry of the Haar measure  $\mu$ , we conclude that

$$\begin{aligned} \int_{O(\mathbb{R}^d)} F_T(x) G(x + Tv_k) d\mu(T) &= \int_H \int_{O(\mathbb{R}^d)} F_{TS}(x) G(x + TSv_k) d\mu(T) dv_H(S) \\ &= \int_{O(\mathbb{R}^d)} \int_H F_T(x) G(x + TSv_k) dv_H(S) d\mu(T), \end{aligned}$$

where we have used that  $F_{TS} = F_T$  for all  $S \in H$ , since by definition,  $Sv_i = v_i$  for all  $1 \leq i \leq k - 1$ . Using this identity, we conclude that the expression we wish to bound is at most

$$\left| \int_{O(\mathbb{R}^d)} \left( \int_{\mathbb{R}^d} \int_H F_T(x) G(x + TSv_k) dv_H(S) dx \right) d\mu(T) \right|. \tag{2}$$

Now we concentrate on the expression inside the parenthesis in (2) for some fixed  $T \in O(\mathbb{R}^d)$ . We claim that, when  $S$  is distributed according to the Haar measure on  $H$ , the variable  $y := TS(v_k/\|v_k\|)$  is uniformly distributed on the unit sphere of the subspace  $TU^\perp$ . This follows from the fact that  $v_k/\|v_k\|$  is on the unit sphere of  $U^\perp$ , and  $H := \text{Stab}^{O(\mathbb{R}^d)}(U)$  is isomorphic<sup>5</sup> to the orthogonal group of  $U^\perp$ . Denoting by  $\sigma_{TU^\perp}^{(d-k+1)}$  the normalized surface measure on the unit sphere of  $TU^\perp$ , we can then write the expression inside the parenthesis in (2) as

$$\begin{aligned} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} F_T(x) G(x + \|v_k\|y) d\sigma_{TU^\perp}^{(d-k+1)}(y) dx \\ = \int_{\mathbb{R}^d} F_T(x) \int_{\mathbb{R}^d} e^{2\pi i x \cdot \xi} \widehat{G}(\xi) \widehat{\sigma}_{TU^\perp}^{(d-k+1)}(-\|v_k\|\xi) d\xi dx \\ = \int_{\mathbb{R}^d} \widehat{F}_T(-\xi) \widehat{G}(\xi) \widehat{\sigma}_{TU^\perp}^{(d-k+1)}(-\|v_k\|\xi) d\xi. \end{aligned}$$

Integrating over  $T \in O(\mathbb{R}^d)$  and applying Cauchy-Schwarz to the inner integral, we conclude that (2) is at most

$$\begin{aligned} \int_{O(\mathbb{R}^d)} \|\widehat{F}_T\|_2 \left( \int_{\mathbb{R}^d} |\widehat{G}(\xi)|^2 |\widehat{\sigma}_{TU^\perp}^{(d-k+1)}(\|v_k\|\xi)|^2 d\xi \right)^{1/2} d\mu(T) \\ = \int_{O(\mathbb{R}^d)} \|F_T\|_2 \left( \int_{\mathbb{R}^d} |\widehat{f}_k(\xi)|^2 |1 - \widehat{Q}_\delta(\xi)|^2 |\widehat{\sigma}_{TU^\perp}^{(d-k+1)}(\|v_k\|\xi)|^2 d\xi \right)^{1/2} d\mu(T), \end{aligned}$$

where we have used Parseval’s Identity and the convolution identity.

Since  $|F_T(x)| \leq |f_1(x + Tv_1)|$  pointwise, it follows that  $\|F_T\| \leq \|f_1\|_2$  for all  $T \in O(\mathbb{R}^d)$ . Using this inequality and applying Cauchy-Schwarz to the outer integral, we see that the expression above is at most

$$\|f_1\|_2 \left( \int_{O(\mathbb{R}^d)} \int_{\mathbb{R}^d} |\widehat{f}_k(\xi)|^2 |1 - \widehat{Q}_\delta(\xi)|^2 |\widehat{\sigma}_{TU^\perp}^{(d-k+1)}(\|v_k\|\xi)|^2 d\xi d\mu(T) \right)^{1/2}. \tag{3}$$

Finally, the double integral in (3) can be bounded using the Fourier estimates given in Lemma 3, as we now show. Divide the integral over  $\mathbb{R}^d$  into two parts, corresponding to the bounded region where  $\|\xi\| \leq (\delta\|v_k\|)^{-1/2}$  and the unbounded region where  $\|\xi\| > (\delta\|v_k\|)^{-1/2}$ . For the bounded region, we

<sup>5</sup>Every orthogonal transformation on  $U^\perp$  can be identified with an element of  $\text{Stab}^{O(\mathbb{R}^d)}(U)$  by tensoring with the identity on  $U$ , with this identification being bijective and measure-preserving.

note that  $|\widehat{\sigma}_{TU^\perp}^{(d-k+1)}(\xi)| \leq 1$  for all  $T \in O(\mathbb{R}^d)$ ,  $\xi \in \mathbb{R}^d$  and use the first inequality in Lemma 3 to obtain

$$\begin{aligned} \int_{O(\mathbb{R}^d)} \int_{\|\xi\| \leq (\delta \|v_k\|)^{-1/2}} |\widehat{f}_k(\xi)|^2 |1 - \widehat{Q}_\delta(\xi)|^2 |\widehat{\sigma}_{TU^\perp}^{(d-k+1)}(\|v_k\|\xi)|^2 d\xi d\mu(T) \\ \leq \int_{O(\mathbb{R}^d)} \int_{\|\xi\| \leq (\delta \|v_k\|)^{-1/2}} |\widehat{f}_k(\xi)|^2 (C_1 \delta^2 \|\xi\|^2)^2 d\xi d\mu(T) \\ \leq C_1^2 \delta^2 \|v_k\|^{-2} \|f_k\|_2^2. \end{aligned}$$

For the unbounded region, we use the simple estimate  $|\widehat{Q}_\delta(\xi)| \leq \|Q_\delta\|_1 = 1$  and the second inequality in Lemma 3 to conclude that

$$\begin{aligned} \int_{O(\mathbb{R}^d)} \int_{\|\xi\| > (\delta \|v_k\|)^{-1/2}} |\widehat{f}_k(\xi)|^2 |1 - \widehat{Q}_\delta(\xi)|^2 |\widehat{\sigma}_{TU^\perp}^{(d-k+1)}(\|v_k\|\xi)|^2 d\xi d\mu(T) \\ \leq \int_{\|\xi\| > (\delta \|v_k\|)^{-1/2}} 4|\widehat{f}_k(\xi)|^2 \int_{O(\mathbb{R}^d)} |\widehat{\sigma}_{TU^\perp}^{(d-k+1)}(\|v_k\|\xi)|^2 d\mu(T) d\xi \\ \leq 4\|f_k\|_2^2 \sup_{\|\xi\| > (\delta \|v_k\|)^{-1/2}} \int_{O(\mathbb{R}^d)} |\widehat{\sigma}_{TU^\perp}^{(d-k+1)}(\|v_k\|\xi)|^2 d\mu(T) \\ \leq 4C_2 (\delta \|v_k\|^{-1})^{(d-k+1)/2} \|f_k\|_2^2. \end{aligned}$$

Summing the bounds obtained for both regions shows that, for  $d \geq k$  and  $0 < \delta \leq 1$ , we can bound expression (3) by

$$(C_1^2 \|v_k\|^{-2} + 4C_2 \|v_k\|^{-(d-k+1)/2})^{1/2} \delta^{1/4} \|f_1\|_2 \|f_k\|_2,$$

and the inequality in the statement of the lemma follows. Since this last bound depends continuously on the positioning of the points of  $P$  (which gives the value of  $\|v_k\|$ ), the claim that the obtained constant  $C_P$  can be made uniform inside a neighborhood of  $P$  also follows.  $\square$

We remark that the proof above is the only place (in the Euclidean setting) where we make explicit use of the assumption that a configuration is admissible. However, as the Counting Lemma will be an essential ingredient of several later results, this requirement will be inherited by them as well.

### 2.2. Continuity properties of the counting function

Given some configuration  $P$  on the space  $\mathbb{R}^d$ , it is sometimes important to understand how much the count of congruent copies of  $P$  on a set  $A \subseteq \mathbb{R}^d$  can change if we perturb the set  $A$  a little. An instance of this problem was already considered in the Counting Lemma, where the perturbation was given by blurring, and it was seen that the counting function  $I_P$  is somewhat robust to small perturbations (in the case of admissible configurations).

Using our telescoping sum trick, it is easy to show that  $I_P$  is also robust to small perturbations measured by the  $L^\infty$  norm; more precisely,  $I_P$  is continuous on  $L^\infty(Q(0, R))$  for any fixed  $R > 0$ . When  $P$  is admissible, we obtain the following significantly stronger continuity property.

**Lemma 5** (Weak\* continuity). *If  $P \subset \mathbb{R}^d$  is an admissible configuration, then for every fixed  $R > 0$ , the function  $I_P$  is weak\* continuous on the unit ball of  $L^\infty(Q(0, R))$ .*

*Proof.* Denote the closed unit ball of  $L^\infty(Q(0, R))$  by  $\mathcal{B}_\infty$ . Since  $\mathcal{B}_\infty$  endowed with the weak\* topology is metrizable (see e.g., [15, Corollary 2.6.20]), it suffices to prove that  $I_P$  is sequentially continuous (i.e., that  $I_P(f_i) \xrightarrow{i \rightarrow \infty} I_P(f)$  whenever  $f_i \xrightarrow{i \rightarrow \infty} f$ ).

Suppose then  $(f_i)_{i \geq 1} \subset \mathcal{B}_\infty$  is a sequence weak\* converging to  $f \in \mathcal{B}_\infty$ . It follows that, for every  $x \in Q(0, R)$  and every  $\delta > 0$ , we have

$$f_i * \mathcal{Q}_\delta(x) = \delta^{-d} \int_{Q(x, \delta)} f_i(y) dy \xrightarrow{i \rightarrow \infty} \delta^{-d} \int_{Q(x, \delta)} f(y) dy = f * \mathcal{Q}_\delta(x).$$

Since  $f * \mathcal{Q}_\delta$  and each  $f_i * \mathcal{Q}_\delta$  are Lipschitz with the same constant (depending only on  $\delta$ , as  $\|f\|_\infty, \|f_i\|_\infty \leq 1$ ) and  $Q(0, R)$  is bounded, this easily implies that

$$\|f_i * \mathcal{Q}_\delta - f * \mathcal{Q}_\delta\|_\infty \rightarrow 0 \text{ as } i \rightarrow \infty.$$

In particular, it follows that  $\lim_{i \rightarrow \infty} I_P(f_i * \mathcal{Q}_\delta) = I_P(f * \mathcal{Q}_\delta)$ .

Since  $P$  is admissible, by the Counting Lemma (Lemma 4), we have

$$|I_P(f * \mathcal{Q}_\delta) - I_P(f)|, |I_P(f_i * \mathcal{Q}_\delta) - I_P(f_i)| \leq C_P \delta^{1/4} R^d \text{ for all } i \geq 1.$$

Choosing  $i_0(\delta) \geq 1$  sufficiently large so that

$$|I_P(f_i * \mathcal{Q}_\delta) - I_P(f * \mathcal{Q}_\delta)| \leq C_P \delta^{1/4} R^d \text{ for all } i \geq i_0(\delta),$$

we conclude that

$$\begin{aligned} |I_P(f) - I_P(f_i)| &\leq |I_P(f) - I_P(f * \mathcal{Q}_\delta)| + |I_P(f * \mathcal{Q}_\delta) - I_P(f_i * \mathcal{Q}_\delta)| \\ &\quad + |I_P(f_i * \mathcal{Q}_\delta) - I_P(f_i)| \\ &\leq 3C_P \delta^{1/4} R^d \text{ for all } i \geq i_0(\delta). \end{aligned}$$

Since  $\delta > 0$  is arbitrary, this implies that  $\lim_{i \rightarrow \infty} I_P(f_i) = I_P(f)$ , as wished. □

We will also need an *equicontinuity* property for the family of counting functions  $P \mapsto I_P(A)$  over all bounded measurable sets  $A \subset \mathbb{R}^d$ . In what follows, we shall write  $\mathcal{B}(P, r) \subset (\mathbb{R}^d)^k$  for the ball of radius  $r$  centered on  $P = \{v_1, \dots, v_k\}$ , where we recall that the distance from  $P$  to  $Q = \{u_1, \dots, u_k\}$  is given by

$$\|P - Q\|_\infty = \min_{\sigma \in \mathfrak{S}_k} \max_{1 \leq i \leq k} \|v_i - u_{\sigma(i)}\|.$$

**Lemma 6** (Equicontinuity). *For every admissible  $P \subset \mathbb{R}^d$  and every  $\varepsilon > 0$ , there is  $\delta > 0$  such that the following holds: if  $P' \in \mathcal{B}(P, \delta)$ , then for all  $R \geq 1$ , we have*

$$|I_{P'}(A) - I_P(A)| \leq \varepsilon R^d \text{ for all measurable } A \subseteq Q(0, R).$$

*Proof.* We will use the fact that the constant  $C_P$  promised in the Counting Lemma can be made uniform inside a small neighborhood of  $P$ ; more precisely, there is  $r > 0$  and a constant  $\tilde{C}_P > 0$  such that

$$|I_{P'}(A) - I_{P'}(A * \mathcal{Q}_\rho)| \leq \tilde{C}_P \rho^{1/4} R^d \text{ for all } \rho \in (0, 1]$$

holds for all  $P' \in \mathcal{B}(P, r)$ ,  $R > 0$  and (measurable)  $A \subseteq Q(0, R)$ .

Fix constants  $R \geq 1$  and  $\delta, \rho \in (0, 1]$  with  $\delta < \rho$ . For any set  $A \subseteq Q(0, R)$  and any points  $x, y \in Q(0, R)$  with  $\|x - y\| \leq \delta$ , we have that

$$\begin{aligned} |A * \mathcal{Q}_\rho(x) - A * \mathcal{Q}_\rho(y)| &= \frac{|\text{vol}(A \cap Q(x, \rho)) - \text{vol}(A \cap Q(y, \rho))|}{\rho^d} \\ &\leq \frac{\text{vol}(Q(x, \rho) \setminus Q(y, \rho))}{\rho^d} \\ &\leq \frac{\rho^d - (\rho - \delta)^d}{\rho^d}. \end{aligned}$$

Noting that  $A * \mathcal{Q}_\rho$  is supported on  $Q(0, R + \rho) \subseteq Q(0, 2R)$ , we conclude from our telescoping sum trick that

$$|I_{P'}(A * \mathcal{Q}_\rho) - I_P(A * \mathcal{Q}_\rho)| \leq k \frac{\rho^d - (\rho - \delta)^d}{\rho^d} (2R)^d$$

whenever  $\|P' - P\|_\infty \leq \delta$ .

Now take  $\rho \in (0, 1]$  small enough so that  $\tilde{C}_P \rho^{1/4} \leq \varepsilon/4$ , and for this value of  $\rho$ , take  $0 < \delta < \min\{r, \rho\}$  small enough so that

$$(\rho - \delta)^d \geq \left(1 - \frac{\varepsilon}{2^{d+1}k}\right) \rho^d.$$

Then, for any configuration  $P' \in \mathcal{B}(P, \delta)$  and any set  $A \subseteq Q(0, R)$ , we obtain

$$\begin{aligned} |I_{P'}(A) - I_P(A)| &\leq |I_{P'}(A) - I_{P'}(A * \mathcal{Q}_\rho)| + |I_{P'}(A * \mathcal{Q}_\rho) - I_P(A * \mathcal{Q}_\rho)| \\ &\quad + |I_P(A * \mathcal{Q}_\rho) - I_P(A)| \\ &\leq \tilde{C}_P \rho^{1/4} R^d + k \frac{\rho^d - (\rho - \delta)^d}{\rho^d} (2R)^d + \tilde{C}_P \rho^{1/4} R^d \\ &\leq \frac{\varepsilon}{4} R^d + k \frac{\varepsilon}{2^{d+1}k} (2R)^d + \frac{\varepsilon}{4} R^d = \varepsilon R^d, \end{aligned}$$

as desired. □

### 2.3. The Supersaturation Theorem

Now we wish to show that geometrical hypergraphs encoding copies of some admissible configuration  $P$  have a nice *supersaturation property*: if a set  $A \subseteq \mathbb{R}^d$  is just slightly denser than the independence density of  $P$ , then it must contain a positive proportion of all congruent copies of  $P$ . This result is quite similar, both formally and in spirit, to an important combinatorial theorem of Erdős and Simonovits [10] in the setting of forbidden graphs and hypergraphs.

**Remark.** The insight that supersaturation results can be used to better study extremal geometrical problems of the kind we are interested in is due to Bukh [3]. He introduced the notion of a ‘supersaturable property’ as any characteristic of measurable sets which satisfies several conditions meant to enable the proof of a supersaturation result; the prototypical and most important example of supersaturable property given in Bukh’s paper is that of avoiding a finite collection of distances. Here, we will obtain similar results in the case of avoiding general admissible configurations, but our method of proof is more analytical in nature and quite different from his.

Using our zero-measure removal lemma (Lemma 2), we can immediately obtain a weak supersaturation property which holds for any  $R > 0$  and any configuration  $P \subset \mathbb{R}^d$ :

(WS) If  $d_{Q(0,R)}(A) > \mathbf{m}_{Q(0,R)}(P)$ , then  $I_P(A) > 0$ .

For our purposes, however, we will need to strengthen this simple property in two ways: first to obtain a uniform lower bound on  $I_P(A)$  which depends only on  $R$  and the slack  $d_{Q(0,R)}(A) - \mathbf{m}_{Q(0,R)}(P)$ , but not on the specific set  $A \subseteq Q(0, R)$ ; and then to make the proportion  $I_P(A \cap Q(0, R))/R^d$  of copies of  $P$  uniform also on the size  $R$  of the cube considered.

The first strengthening can be obtained from (WS) by a compactness argument, using the fact that the counting function of admissible configurations is weak\* continuous.

**Lemma 7** (Weak supersaturation). *Let  $P \subset \mathbb{R}^d$  be an admissible configuration. For every  $R > 0$  and  $\varepsilon > 0$ , there exists  $c_0 > 0$  such that the following holds: whenever  $A \subseteq Q(0, R)$  satisfies  $d_{Q(0,R)}(A) \geq \mathbf{m}_{Q(0,R)}(P) + \varepsilon$ , we have  $I_P(A) \geq c_0$ .*

*Proof.* Suppose, for contradiction, that the result is false. Then, there exist  $\varepsilon > 0, R > 0$  and a sequence  $(A_i)_{i \geq 1}$  of subsets of  $Q(0, R)$ , each of density at least  $\mathbf{m}_{Q(0,R)}(P) + \varepsilon$ , which satisfy  $\lim_{i \rightarrow \infty} I_P(A_i) = 0$ .

The unit ball  $\mathcal{B}_\infty$  of  $L^\infty(Q(0, R))$  is weak\* compact by the Banach-Alaoglu Theorem, and it is also metrizable in this topology (see [15, Chapter 2.6]). By possibly restricting to a subsequence, we may then assume that  $(A_i)_{i \geq 1}$  converges in the weak\* topology of  $L^\infty(Q(0, R))$ ; let us denote its limit by  $A \in \mathcal{B}_\infty$ . It is clear that  $0 \leq A \leq 1$  almost everywhere, and

$$\frac{1}{R^d} \int_{Q(0,R)} A(x) \, dx = \lim_{i \rightarrow \infty} \frac{1}{R^d} \int_{Q(0,R)} A_i(x) \, dx \geq \mathbf{m}_{Q(0,R)}(P) + \varepsilon.$$

By weak\* continuity of  $I_P$  (Lemma 5), we also have  $I_P(A) = \lim_{i \rightarrow \infty} I_P(A_i) = 0$ .

Now, let  $B := \{x \in Q(0, R) : A(x) \geq \varepsilon\}$ . Since

$$\varepsilon B(x) \leq A(x) < \varepsilon + B(x) \quad \text{for a.e. } x \in Q(0, R),$$

we conclude that  $I_P(B) \leq \varepsilon^{-k} I_P(A) = 0$  and

$$d_{Q(0,R)}(B) > \frac{1}{R^d} \int_{Q(0,R)} A(x) \, d\sigma(x) - \varepsilon \geq \mathbf{m}_{Q(0,R)}(P).$$

But this set  $B$  contradicts (WS) (or Lemma 2), finishing the proof. □

Our desired supersaturation result now follows from a simple averaging argument.

**Theorem 3** (Supersaturation Theorem). *Let  $P \subset \mathbb{R}^d$  be an admissible configuration and let  $\varepsilon > 0$ . There exist constants  $c > 0$  and  $R_0 > 0$  such that the following holds for all  $R \geq R_0$ : if  $A \subseteq Q(0, R)$  satisfies*

$$d_{Q(0,R)}(A) \geq \mathbf{m}_{Q(0,R)}(P) + \varepsilon,$$

*then  $I_P(A) \geq cR^d$ .*

*Proof.* Take  $R_1 > 0$  large enough so that  $\mathbf{m}_{Q(0,R_1)}(P) \leq \mathbf{m}_{\mathbb{R}^d}(P) + \varepsilon/4$  (see Lemma 1). We will show that the conclusion of the theorem holds for  $R_0 = 4dR_1/\varepsilon$  and some constant  $c > 0$  to be chosen later.

Suppose  $R \geq 4dR_1/\varepsilon$ , and let  $A \subseteq Q(0, R)$  be a set of density

$$d_{Q(0,R)}(A) \geq \mathbf{m}_{Q(0,R)}(P) + \varepsilon.$$

Since  $\mathbf{m}_{Q(0,R_1)}(P) \leq \mathbf{m}_{\mathbb{R}^d}(P) + \varepsilon/4 \leq \mathbf{m}_{Q(0,R)}(P) + \varepsilon/4$ , we have that

$$\text{vol}(A) = \text{vol}(A \cap Q(0, R)) \geq (\mathbf{m}_{Q(0,R_1)}(P) + 3\varepsilon/4)R^d.$$

Let  $K := \lfloor R/R_1 \rfloor$ , and note that

$$K^d R_1^d > \left(1 - \frac{R_1}{R}\right)^d R^d \geq \left(1 - \frac{dR_1}{R}\right) R^d.$$



By our assumption on  $R$ , we conclude that  $K^d R_1^d \geq (1 - \varepsilon/4)R^d$ , and thus

$$\begin{aligned} \text{vol}(A \cap Q(0, KR_1)) &\geq \text{vol}(A) - \text{vol}(Q(0, R) \setminus Q(0, KR_1)) \\ &\geq (\mathbf{m}_{Q(0, R_1)}(P) + \varepsilon/2)K^d R_1^d. \end{aligned}$$

Partitioning the cube  $Q(0, KR_1)$  into  $K^d$  cubes of side length  $R_1$ , by averaging, we conclude that at least  $\varepsilon K^d/4$  of these cubes  $Q(x, R_1)$  satisfy

$$\text{vol}(A \cap Q(x, R_1)) \geq (\mathbf{m}_{Q(0, R_1)}(P) + \varepsilon/4)R_1^d. \tag{4}$$

By Lemma 7, there is some  $c_0 > 0$  (depending on  $R_1$  and  $\varepsilon$  but not on  $A$  or  $R$ ) such that  $I_P(A \cap Q(x, R_1)) \geq c_0$  holds for each one of the cubes in the partition satisfying (4); summing up all these values, we conclude that

$$I_P(A) \geq \frac{\varepsilon K^d}{4} c_0 > \frac{\varepsilon c_0}{4} \left(\frac{R}{R_1} - 1\right)^d \geq \frac{\varepsilon c_0}{4} \frac{1}{(2R_1)^d} R^d,$$

finishing the proof for  $c = \varepsilon c_0 / (2^{d+2} R_1^d)$ . □

**Remark.** The arguments used in the proofs of Lemma 7 and Theorem 3 easily extend to the case of several configurations  $P_1, \dots, P_n \subset \mathbb{R}^d$ , with only minor and notational modifications. In the case of the Supersaturation Theorem, one concludes that  $I_{P_i}(A) \geq c(\varepsilon)R^d$  holds for some  $1 \leq i \leq n$  whenever the density condition  $d_{Q(0, R)}(A) \geq \mathbf{m}_{Q(0, R)}(P_1, \dots, P_n) + \varepsilon$  is satisfied (assuming  $R$  is large enough and all the configurations  $P_i$  are admissible).

Following Bukh [3], for each  $\delta > 0$  and  $\gamma > 0$ , we define the *zooming-out operator*  $\mathcal{Z}_\delta(\gamma)$  as the map which takes a measurable set  $A \subseteq \mathbb{R}^d$  to the set

$$\mathcal{Z}_\delta(\gamma)[A] := \{x \in \mathbb{R}^d : d_{Q(x, \delta)}(A) \geq \gamma\}.$$

Intuitively,  $\mathcal{Z}_\delta(\gamma)[A]$  represents the points where  $A$  is not too sparse at scale  $\delta$ .

Using the Supersaturation Theorem together with the Counting Lemma, we can now show that the existence of copies of  $P$  in a set  $A$  follows also from the weaker assumption that its zoomed-out version  $\mathcal{Z}_\delta(\gamma)[A]$  has density higher than  $\mathbf{m}_{\mathbb{R}^d}(P)$  (rather than  $A$  itself having this same density). This property will be important for us later on.

**Corollary 1.** *Given an admissible configuration  $P \subset \mathbb{R}^d$  and  $\varepsilon > 0$ , there exists  $\delta_0 > 0$  such that the following holds for all  $\delta \leq \delta_0$ : if  $A \subseteq \mathbb{R}^d$  satisfies*

$$\overline{d}(\mathcal{Z}_\delta(\varepsilon)[A]) \geq \mathbf{m}_{\mathbb{R}^d}(P) + \varepsilon,$$

*then  $A$  contains a congruent copy of  $P$ .*

*Proof.* Let  $R_0, c > 0$  be the constants promised in the Supersaturation Theorem applied to  $P$  and with  $\varepsilon$  substituted by  $\varepsilon/3$ . Up to substituting  $R_0$  by some larger constant, we may also assume that  $\mathbf{m}_{Q(0, R)}(P) \leq \mathbf{m}_{\mathbb{R}^d}(P) + \varepsilon/3$  for all  $R \geq R_0$  (see Lemma 1).

Suppose  $A \subseteq \mathbb{R}^d$  satisfies  $\overline{d}(\mathcal{Z}_\delta(\varepsilon)[A]) \geq \mathbf{m}_{\mathbb{R}^d}(P) + \varepsilon$  for some  $0 < \delta \leq 1$ . Since

$$\begin{aligned} \limsup_{R \rightarrow \infty} d_{Q(0, R)}(\mathcal{Z}_\delta(\varepsilon)[A \cap Q(0, R)]) &= \limsup_{R \rightarrow \infty} d_{Q(0, R)}(\mathcal{Z}_\delta(\varepsilon)[A]) \\ &= \overline{d}(\mathcal{Z}_\delta(\varepsilon)[A]) \geq \mathbf{m}_{\mathbb{R}^d}(P) + \varepsilon, \end{aligned}$$

there must exist some  $R \geq R_0$  such that

$$d_{Q(0, R)}(\mathcal{Z}_\delta(\varepsilon)[A \cap Q(0, R)]) \geq \mathbf{m}_{\mathbb{R}^d}(P) + 2\varepsilon/3.$$

Denoting  $A' := A \cap Q(0, R)$ , we may then assume that  $A' \subseteq Q(0, R)$  satisfies

$$d_{Q(0,R)}(\mathcal{Z}_\delta(\varepsilon)[A']) \geq \mathbf{m}_{Q(0,R)}(P) + \varepsilon/3 \tag{5}$$

for some  $R \geq R_0$ , and wish to show that  $A'$  (and hence  $A$ ) contains a copy of  $P$  if  $\delta > 0$  is small enough depending on  $P$  and  $\varepsilon$ .

By the Supersaturation Theorem, inequality (5) implies that  $I_P(\mathcal{Z}_\delta(\varepsilon)[A']) \geq cR^d$ . Since  $A' * Q_\delta(x) \geq \varepsilon \cdot \mathcal{Z}_\delta(\varepsilon)[A'](x)$  for all  $x \in \mathbb{R}^d$ , we obtain from the Counting Lemma that

$$\begin{aligned} I_P(A') &\geq I_P(A' * Q_\delta) - C_P \delta^{1/4} R^d \geq \varepsilon^{|P|} I_P(\mathcal{Z}_\delta(\varepsilon)[A']) - C_P \delta^{1/4} R^d \\ &\geq (\varepsilon^{|P|} c - C_P \delta^{1/4}) R^d. \end{aligned}$$

Taking  $\delta > 0$  small enough for this last expression to be positive, we conclude that  $I_P(A') > 0$ , and so  $A'$  contains a copy of  $P$  as wished. □

### 2.4. Results on the independence density

We are finally in a position to properly study the independence density parameter for a family of configurations in Euclidean space.

We start by proving a simple lower bound on the independence density of several distinct configurations; this result and the argument we use to prove it are originally due to Bukh [3].

**Lemma 8** (Supermultiplicativity). *For all  $n \geq 1$  and all configurations  $P_1, \dots, P_n \subset \mathbb{R}^d$ , we have that*

$$\mathbf{m}_{\mathbb{R}^d}(P_1, P_2, \dots, P_n) \geq \prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(P_i).$$

*Proof.* Fix  $\varepsilon > 0$  and choose  $R$  large enough so that

$$\min_{1 \leq i \leq n} (R - \text{diam } P_i)^d \geq (1 - \varepsilon) R^d.$$

For each  $1 \leq i \leq n$ , let  $A_i \subseteq Q(0, R - \text{diam } P_i)$  be a set which avoids  $P_i$  and satisfies  $d_{Q(0, R - \text{diam } P_i)}(A_i) > \mathbf{m}_{\mathbb{R}^d}(P_i) - \varepsilon$  (this is possible by Lemma 1). We then construct the  $R$ -periodic set  $A'_i := A_i + R\mathbb{Z}^d$ , which also avoids  $P_i$  and has density

$$d(A'_i) = \frac{(R - \text{diam } P_i)^d}{R^d} d_{Q(0, R - \text{diam } P_i)}(A_i) > \mathbf{m}_{\mathbb{R}^d}(P_i) - 2\varepsilon.$$

Since each set  $A'_i$  is periodic with the same fundamental domain  $Q(0, R)$ , it follows that the average of  $d(\bigcap_{i=1}^n (x_i + A'_i))$  over independent translates  $x_1, \dots, x_n \in Q(0, R)$  is equal to  $\prod_{i=1}^n d(A'_i)$ . There must then exist some  $x_1, \dots, x_n \in Q(0, R)$  for which

$$d\left(\bigcap_{i=1}^n (x_i + A'_i)\right) \geq \prod_{i=1}^n d(A'_i) > \prod_{i=1}^n (\mathbf{m}_{\mathbb{R}^d}(P_i) - 2\varepsilon).$$

Since  $\bigcap_{i=1}^n (x_i + A'_i)$  avoids each of the configurations  $P_i$  and  $\varepsilon > 0$  was arbitrary, the desired lower bound follows. □

Intuitively, one may regard  $\mathbf{m}_{\mathbb{R}^d}(P_1, P_2, \dots, P_n)$  being close to  $\prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(P_i)$  as some sort of *independence* or *lack of correlation* between the  $n$  constraints of forbidding each configuration  $P_i$ ; in this case, there is no better way to choose a set avoiding all of these configurations than simply intersecting optimal  $P_i$ -avoiding sets for each  $i$  (after suitably translating them). One might then expect

this to happen if the sizes of each  $P_i$  are very different from each other, so that each constraint will be relevant in different and largely independent scales.

Our next result shows this is indeed the case whenever the configurations considered are all admissible. (A theorem of Graham [13] implies this is *not* necessarily true if the configurations considered are non-admissible; see Section 4 for a discussion.) The proof we present here is based on Bukh’s arguments for supersaturable properties and generalizes his result from two-point configurations to general admissible configurations.

**Theorem 4** (Asymptotic independence). *If  $P_1, P_2, \dots, P_n \subset \mathbb{R}^d$  are admissible configurations, then*

$$\mathbf{m}_{\mathbb{R}^d}(t_1P_1, t_2P_2, \dots, t_nP_n) \rightarrow \prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(P_i)$$

as the ratios  $t_2/t_1, t_3/t_2, \dots, t_n/t_{n-1}$  tend to infinity.

*Proof.* We have already seen that

$$\mathbf{m}_{\mathbb{R}^d}(t_1P_1, t_2P_2, \dots, t_nP_n) \geq \prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(t_iP_i) = \prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(P_i)$$

always holds, so it suffices to show that  $\mathbf{m}_{\mathbb{R}^d}(t_1P_1, t_2P_2, \dots, t_nP_n)$  is no larger than  $\prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(P_i) + \varepsilon$  whenever  $\varepsilon > 0$  and the ratios between consecutive scales  $t_i$  are large enough. We shall proceed by induction, with the case  $n = 1$  being trivial.

Let  $n \geq 2$  and suppose the theorem holds for configurations  $P_1, \dots, P_{n-1}$ . Fix  $0 < \varepsilon \leq 1$  and let  $t_1, \dots, t_{n-1} > 0$  be dilation parameters for which

$$\mathbf{m}_{\mathbb{R}^d}(t_1P_1, \dots, t_{n-1}P_{n-1}) \leq \prod_{i=1}^{n-1} \mathbf{m}_{\mathbb{R}^d}(P_i) + \varepsilon;$$

now, take  $R_0 > 0$  large enough so that

$$\mathbf{m}_{Q(0,R)}(t_1P_1, \dots, t_{n-1}P_{n-1}) \leq \mathbf{m}_{\mathbb{R}^d}(t_1P_1, \dots, t_{n-1}P_{n-1}) + \varepsilon$$

holds for all  $R \geq R_0$  (this quantity exists by Lemma 1).

If  $A \subseteq \mathbb{R}^d$  is a measurable set avoiding  $t_1P_1, \dots, t_{n-1}P_{n-1}$ , then clearly

$$d_{Q(x,R)}(A) \leq \mathbf{m}_{Q(0,R)}(t_1P_1, \dots, t_{n-1}P_{n-1}) \quad \text{for all } x \in \mathbb{R}^d, R > 0. \tag{6}$$

Moreover, if  $A$  also avoids  $t_nP_n$  for some  $t_n > 0$ , then  $A/t_n$  avoids  $P_n$  and so by Corollary 1, there is some  $\delta_0 > 0$  (depending only on  $P_n$  and  $\varepsilon$ ) for which

$$\bar{d}(\mathcal{Z}_\delta(\varepsilon)[A/t_n]) \leq \mathbf{m}_{\mathbb{R}^d}(P_n) + \varepsilon \quad \text{for all } \delta \leq \delta_0. \tag{7}$$

Suppose now that  $t_n \geq R_0/\delta_0$  and let  $A \subseteq \mathbb{R}^d$  be any measurable set avoiding  $t_1P_1, \dots, t_nP_n$ . We conclude from (6) that

$$\begin{aligned} d_{Q(x, t_n \delta_0)}(A) &\leq \mathbf{m}_{Q(0, t_n \delta_0)}(t_1P_1, \dots, t_{n-1}P_{n-1}) \\ &\leq \mathbf{m}_{\mathbb{R}^d}(t_1P_1, \dots, t_{n-1}P_{n-1}) + \varepsilon \\ &\leq \prod_{i=1}^{n-1} \mathbf{m}_{\mathbb{R}^d}(P_i) + 2\varepsilon \end{aligned}$$

holds for all  $x \in \mathbb{R}^d$ , and from (7), we have

$$\bar{d}(\mathcal{Z}_{t_n \delta_0}(\varepsilon)[A]) = \bar{d}(\mathcal{Z}_{\delta_0}(\varepsilon)[A/t_n]) \leq \mathbf{m}_{\mathbb{R}^d}(P_n) + \varepsilon.$$

This means that the density of  $A$  inside cubes  $Q(x, t_n \delta_0)$  of side length  $t_n \delta_0$  is at most  $\varepsilon$  (when  $x \notin \mathcal{Z}_{t_n \delta_0}(\varepsilon)[A]$ ) except at a set of upper density at most  $\mathbf{m}_{\mathbb{R}^d}(P_n) + \varepsilon$ , when it is instead no more than  $\prod_{i=1}^{n-1} \mathbf{m}_{\mathbb{R}^d}(P_i) + 2\varepsilon$ . Taking averages, we conclude that

$$\bar{d}(A) \leq \varepsilon + (\mathbf{m}_{\mathbb{R}^d}(P_n) + \varepsilon) \left( \prod_{i=1}^{n-1} \mathbf{m}_{\mathbb{R}^d}(P_i) + 2\varepsilon \right) \leq \prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(P_i) + 6\varepsilon$$

(where we used that  $0 < \varepsilon \leq 1$ ). This inequality finishes the proof. □

As an immediate corollary of the last theorem, we conclude that

$$\mathbf{m}_{\mathbb{R}^d}(t_1 P, t_2 P, \dots, t_n P) \rightarrow \mathbf{m}_{\mathbb{R}^d}(P)^n$$

as  $t_2/t_1, t_3/t_2, \dots, t_n/t_{n-1} \rightarrow \infty$  whenever  $P \subset \mathbb{R}^d$  is admissible; this answers our question (Q1) in the case of admissible configurations. Let us now show how this result easily implies Bourgain’s Theorem given in the Introduction.

*Proof of Theorem 2.* Suppose  $A \subset \mathbb{R}^d$  is a measurable set not satisfying the conclusion of the theorem; thus, there is a sequence  $(t_j)_{j \geq 1}$  tending to infinity such that  $A$  does not contain a copy of any  $t_j P$ . This implies that  $\bar{d}(A) \leq \mathbf{m}_{\mathbb{R}^d}(t_1 P, t_2 P, \dots, t_n P)$  for all  $n \in \mathbb{N}$ . By taking a suitably fast-growing subsequence, we may then use Theorem 4 to obtain (say)  $\bar{d}(A) \leq 2\mathbf{m}_{\mathbb{R}^d}(P)^n$  for any fixed  $n \geq 1$ . Since  $\mathbf{m}_{\mathbb{R}^d}(P) < 1$ ,<sup>6</sup> this implies that  $\bar{d}(A) = 0$ , as wished. □

Going back to our study of the independence density for multiple configurations, we will now consider the opposite situation of what we have seen before: when the constraints of forbidding each individual configuration are so strongly correlated as to be essentially redundant. One might expect this to be the case, for instance, when we are forbidding very close dilates of a given configuration  $P$ .

We will show that this intuition is indeed correct, whether or not the configuration considered is admissible, and the proof is much simpler than in the case of very distant dilates of  $P$  (in particular, not needing the results from earlier sections).

**Lemma 9** (Asymptotic redundancy). *For any configuration  $P \subset \mathbb{R}^d$ , we have that*

$$\mathbf{m}_{\mathbb{R}^d}(t_1 P, t_2 P, \dots, t_n P) \rightarrow \mathbf{m}_{\mathbb{R}^d}(P)$$

as  $t_2/t_1, t_3/t_2, \dots, t_n/t_{n-1} \rightarrow 1$ .

*Proof.* Assume, by dilation invariance, that  $t_1 = 1$ , and note that it suffices to show the convergence above with  $\mathbf{m}_{\mathbb{R}^d}$  replaced by  $\mathbf{m}_{Q(0,R)}$  for every fixed  $R > 0$ . We will then fix an arbitrary  $R > 0$  and prove that  $\mathbf{m}_{Q(0,R)}(P, t_2 P, \dots, t_n P) \rightarrow \mathbf{m}_{Q(0,R)}(P)$  as  $t_2, t_3, \dots, t_n \rightarrow 1$ .

Let  $(v_1, v_2, \dots, v_k)$  be an ordering of the points of  $P$ , and consider the continuous function  $g_P : (\mathbb{R}^d)^k \times \mathcal{O}(\mathbb{R}^d) \rightarrow \mathbb{R}$  given by

$$g_P(x_1, \dots, x_k, T) := \sum_{j=2}^k \|(x_j - x_1) - T(v_j - v_1)\|.$$

Note that  $\min_{T \in \mathcal{O}(\mathbb{R}^d)} g_P(x_1, \dots, x_k, T) = 0$  if and only if  $(x_1, \dots, x_k)$  is congruent to  $(v_1, \dots, v_k)$ .

<sup>6</sup>An easy averaging argument shows that  $\mathbf{m}_{\mathbb{R}^d}(P) \leq 1 - 1/|P|$ ; see Lemma 10 for a proof of this inequality in the spherical setting.

Fix some  $\varepsilon > 0$  and let  $A \subset Q(0, R)$  be a measurable set which avoids  $P$  and has density  $d_{Q(0,R)}(A) \geq \mathbf{m}_{Q(0,R)}(P) - \varepsilon$ . By inner regularity, we know there exists a compact set  $\tilde{A} \subseteq A$  with  $d_{Q(0,R)}(\tilde{A}) \geq \mathbf{m}_{Q(0,R)}(P) - 2\varepsilon$ . Denote by  $\gamma$  the minimum of the continuous function  $g_P$  on the compact set  $\tilde{A}^k \times O(\mathbb{R}^d)$ ; since  $\tilde{A}$  avoids  $P$ , it follows that  $\gamma > 0$ .

We will now prove that  $\tilde{A}$  also avoids  $tP$  whenever  $t$  is sufficiently close to 1, say when  $|t - 1| < \gamma/(k \cdot \text{diam } P)$ . Indeed, for all  $x_1, \dots, x_k \in \tilde{A}$  and all  $T \in O(\mathbb{R}^d)$ , by the triangle inequality, we have that

$$\begin{aligned} \sum_{j=2}^k \|(x_j - x_1) - T(tv_j - tv_1)\| &\geq \sum_{j=2}^k \left( \|(x_j - x_1) - T(v_j - v_1)\| - |t - 1| \|v_j - v_1\| \right) \\ &> \sum_{j=2}^k \|(x_j - x_1) - T(v_j - v_1)\| - k \cdot |t - 1| \text{diam } P \\ &\geq \gamma - k \cdot |t - 1| \text{diam } P, \end{aligned}$$

which is positive if  $|t - 1| < \gamma/(k \cdot \text{diam } P)$ . In particular, we see that

$$\mathbf{m}_{Q(0,R)}(P, t_2P, \dots, t_nP) \geq d_{Q(0,R)}(\tilde{A}) \geq \mathbf{m}_{Q(0,R)}(P) - 2\varepsilon$$

whenever  $|t_j - 1| < \gamma/(k \cdot \text{diam } P)$  for  $2 \leq j \leq n$ . Since we clearly have that  $\mathbf{m}_{Q(0,R)}(P, t_2P, \dots, t_nP) \leq \mathbf{m}_{Q(0,R)}(P)$ , the result follows.  $\square$

The proof of this last result actually implies a somewhat stronger and more technical property of the independence density, namely that every configuration  $P$  where  $\mathbf{m}_{\mathbb{R}^d}$  is *discontinuous* must be a local minimum across the ‘discontinuity barrier’; more formally, we have that  $\inf_{P' \in \mathcal{B}(P, \delta)} \mathbf{m}_{\mathbb{R}^d}(P') \rightarrow \mathbf{m}_{\mathbb{R}^d}(P)$  as  $\delta \rightarrow 0$ , where  $\mathcal{B}(P, \delta)$  is the ball of radius  $\delta$  centered on  $P$  (the details of the proof are given below). If the configuration  $P$  is admissible, then we can also prove the corresponding limit for  $\sup_{P' \in \mathcal{B}(P, \delta)} \mathbf{m}_{\mathbb{R}^d}(P')$  and conclude that  $\mathbf{m}_{\mathbb{R}^d}$  is, in fact, continuous at this point. This is done in the next theorem:

**Theorem 5** (Continuity of the independence density). *For every  $n \geq 1$ , the function  $(P_1, \dots, P_n) \mapsto \mathbf{m}_{\mathbb{R}^d}(P_1, \dots, P_n)$  is continuous on the set of  $n$  admissible configurations in  $\mathbb{R}^d$ .*

*Proof.* For the sake of better readability, we will prove the result in the case of only one forbidden configuration; the  $n$ -variable version easily follows from the same argument. Fix some  $\varepsilon > 0$  and let  $R_1 \geq 1$  be large enough so that

$$\mathbf{m}_{\mathbb{R}^d}(P') \leq \mathbf{m}_{Q(0,R)}(P') \leq \mathbf{m}_{\mathbb{R}^d}(P') + \varepsilon$$

holds for all  $P' \in \mathcal{B}(P, 1)$  and all  $R \geq R_1$  (this value exists by Lemma 1).

Let  $R \geq R_1$  and let  $A \subset Q(0, R)$  be a compact  $P$ -avoiding set with density

$$d_{Q(0,R)}(A) \geq \mathbf{m}_{Q(0,R)}(P) - \varepsilon.$$

Proceeding exactly as we did in the proof of the last lemma, we conclude that  $A$  also avoids all  $P'$  close enough to  $P$ ; for all such configurations, we then have

$$\mathbf{m}_{Q(0,R)}(P') \geq d_{Q(0,R)}(A) \geq \mathbf{m}_{Q(0,R)}(P) - \varepsilon \geq \mathbf{m}_{\mathbb{R}^d}(P) - \varepsilon.$$

Since  $\mathbf{m}_{Q(0,R)}(P') \leq \mathbf{m}_{\mathbb{R}^d}(P') + \varepsilon$  whenever  $P' \in \mathcal{B}(P, 1)$ , this implies that  $\mathbf{m}_{\mathbb{R}^d}(P') \geq \mathbf{m}_{\mathbb{R}^d}(P) - 2\varepsilon$  for all  $P'$  close enough to  $P$ .

Now, we suppose that  $P$  is admissible, and let  $R_0, c > 0$  be the constants promised by the Supersaturation Theorem (Theorem 3). Let  $R \geq \max\{R_0, R_1\}$ . By equicontinuity (Lemma 6), there is some  $\delta > 0$

for which the inequality

$$|I_{P'}(A) - I_P(A)| < cR^d \quad \text{for all } P' \in \mathcal{B}(P, \delta) \tag{8}$$

holds whenever  $A \subset Q(0, R)$  is a measurable set; fix such a value of  $\delta$ .

If  $P' \in \mathcal{B}(P, \delta)$  and  $A \subset Q(0, R)$  is a measurable set avoiding  $P'$ , we conclude from inequality (8) that  $I_{P'}(A) < cR^d$ . By the Supersaturation Theorem, this implies that  $d_{Q(0,R)}(A) < \mathbf{m}_{Q(0,R)}(P) + \varepsilon$ , and thus (by optimizing over  $A$ ), we conclude that  $\mathbf{m}_{Q(0,R)}(P') \leq \mathbf{m}_{Q(0,R)}(P) + \varepsilon$ . It follows that

$$\mathbf{m}_{\mathbb{R}^d}(P') \leq \mathbf{m}_{Q(0,R)}(P') \leq \mathbf{m}_{Q(0,R)}(P) + \varepsilon \leq \mathbf{m}_{\mathbb{R}^d}(P) + 2\varepsilon$$

whenever  $P' \in \mathcal{B}(P, \delta)$ , finishing the proof. □

These last results can now be combined in a very simple way to give an (almost complete) answer to question (Q2), when restricted to admissible configurations. Let us denote by  $\mathcal{M}_n(P)$  the set of all possible independence densities one can obtain by forbidding  $n$  distinct dilates of a configuration  $P$ ; that is,

$$\mathcal{M}_n(P) := \{ \mathbf{m}_{\mathbb{R}^d}(t_1P, t_2P, \dots, t_nP) : 0 < t_1 < t_2 < \dots < t_n < \infty \}.$$

Recall that (Q2) asked for an explicit description of this set  $\mathcal{M}_n(P)$ .

**Theorem 6** (Forbidding multiple dilates). *If  $P \subset \mathbb{R}^d$  is admissible, then*

$$(\mathbf{m}_{\mathbb{R}^d}(P)^n, \mathbf{m}_{\mathbb{R}^d}(P)) \subseteq \mathcal{M}_n(P) \subseteq [\mathbf{m}_{\mathbb{R}^d}(P)^n, \mathbf{m}_{\mathbb{R}^d}(P)].$$

*Proof.* It is clear that  $\mathbf{m}_{\mathbb{R}^d}(t_1P, t_2P, \dots, t_nP) \leq \mathbf{m}_{\mathbb{R}^d}(t_1P) = \mathbf{m}_{\mathbb{R}^d}(P)$  always holds, and we saw in Lemma 8 that

$$\mathbf{m}_{\mathbb{R}^d}(t_1P, t_2P, \dots, t_nP) \geq \prod_{i=1}^n \mathbf{m}_{\mathbb{R}^d}(t_iP) = \mathbf{m}_{\mathbb{R}^d}(P)^n.$$

Moreover, Lemma 9 implies that  $\mathbf{m}_{\mathbb{R}^d}(P)$  is an accumulation point of the set  $\mathcal{M}_n(P)$ , and (since  $P$  is admissible) Theorem 4 implies the same about  $\mathbf{m}_{\mathbb{R}^d}(P)^n$ . The result follows from continuity of the function

$$(t_1, t_2, \dots, t_n) \mapsto \mathbf{m}_{\mathbb{R}^d}(t_1P, t_2P, \dots, t_nP),$$

which is an immediate consequence of Theorem 5. □

As our final result in the Euclidean setting, we will show the existence of extremizer sets which avoid admissible configurations. This generalizes a result of Bukh (see Corollary 13 in [3]) from forbidden distances to higher-order configurations.

**Theorem 7** (Existence of extremizers). *If  $P \subset \mathbb{R}^d$  is admissible, then there exists a  $P$ -avoiding measurable set  $A \subseteq \mathbb{R}^d$  with density  $d(A) = \mathbf{m}_{\mathbb{R}^d}(P)$ .*

*Proof.* For each integer  $i \geq 1$ , let  $A_i \subseteq Q(0, i)$  be a  $P$ -avoiding set with density  $d_{Q(0,i)}(A_i) \geq \mathbf{m}_{Q(0,i)}(P) - 2^{-i}$ . Denote the unit ball of  $L^\infty(\mathbb{R}^d)$  by  $\mathcal{B}_\infty$ ; by the Banach-Alaoglu Theorem,  $\mathcal{B}_\infty$  is weak\* compact. By restricting to a subsequence if necessary, we may then assume that  $(A_i)_{i \geq 1}$  converges to some element  $\tilde{A} \in \mathcal{B}_\infty$  in the weak\* topology of  $L^\infty(\mathbb{R}^d)$ . Denote by  $A := \text{supp } \tilde{A}$  the support of  $\tilde{A}$ .<sup>7</sup>

We will first prove that  $I_P(A) = 0$ . Fix some  $R > 0$  and (for notational convenience) denote the indicator function of the cube  $Q(0, R)$  by  $\chi_R$ . Writing  $\chi_R A_i$  for the pointwise product of  $\chi_R$  and the

<sup>7</sup>Strictly speaking,  $\tilde{A}$  is an equivalence class of functions, not a specific function. More formally, our set  $A$  is the support of an (arbitrary) representative of this class, but since this choice of representative makes no difference to our argument, one can ignore this technicality.

indicator function of  $A_i$ , one easily sees from the definition that  $(\chi_{RA_i})_{i \geq 1}$  converges to  $\chi_{R\tilde{A}}$  in the weak\* topology of  $L^\infty(Q(0, R))$ . As  $P$  is admissible, the counting function  $I_P$  is weak\* continuous (by Lemma 5) and thus,  $I_P(\chi_{R\tilde{A}}) = \lim_{i \rightarrow \infty} I_P(\chi_{RA_i}) = 0$ . We now proceed as in the proof of Lemma 7 to show that  $I_P(\text{supp } \chi_{R\tilde{A}}) = 0$  as well: first, approximate  $I_P(\text{supp } \chi_{R\tilde{A}})$  by  $I_P(B_\varepsilon)$ , where  $B_\varepsilon := \{x \in Q(0, R) : \tilde{A}(x) \geq \varepsilon\}$  and  $\varepsilon > 0$  is a sufficiently small constant (depending on  $R$ ), and then note that  $I_P(B_\varepsilon) \leq \varepsilon^{-|P|} I_P(\chi_{RA}) = 0$  for all  $\varepsilon > 0$ . Since  $\text{supp } \chi_{R\tilde{A}} = A \cap Q(0, R)$  up to zero-measure sets and  $R > 0$  is arbitrary, we conclude that  $I_P(A) = 0$  as wished.

Next we prove that  $d(A) = \mathbf{m}_{\mathbb{R}^d}(P)$ . Since  $I_P(A) = 0$ , it follows from Lemma 2 that  $\bar{d}(A) \leq \mathbf{m}_{\mathbb{R}^d}(P)$ , and so it suffices to show that

$$\liminf_{R \rightarrow \infty} d_{Q(0,R)}(A) \geq \mathbf{m}_{\mathbb{R}^d}(P). \tag{9}$$

Fix some arbitrary  $\varepsilon > 0$  and take  $R_0 \geq 2$  large enough so that  $(R_0 + 2\text{diam } P)^d < (1 + \varepsilon/4)R_0^d$ . For any given  $R \geq R_0$ , take a  $P$ -avoiding set  $B_R \subseteq Q(0, R)$  with

$$d_{Q(0,R)}(B_R) > \mathbf{m}_{Q(0,R)}(P) - \varepsilon/4 \geq \mathbf{m}_{\mathbb{R}^d}(P) - \varepsilon/4.$$

For all  $i \geq R$ , define  $A'_i := B_R \cup (A_i \setminus Q(0, R + 2\text{diam } P))$ ; note that  $A'_i$  avoids  $P$  and

$$\begin{aligned} \text{vol}(A'_i) &= \text{vol}(A_i) - \text{vol}(A_i \cap (Q(0, R + 2\text{diam } P) \setminus Q(0, R))) \\ &\quad - \text{vol}(A_i \cap Q(0, R)) + \text{vol}(B_R) \\ &\geq \text{vol}(A_i) - ((R + 2\text{diam } P)^d - R^d) + \text{vol}(B_R) - \text{vol}(A \cap Q(0, R)) \\ &\quad + \text{vol}(A \cap Q(0, R)) - \text{vol}(A_i \cap Q(0, R)) \\ &\geq (\mathbf{m}_{Q(0,i)}(P) - 2^{-i})i^d - \frac{\varepsilon R^d}{4} + (d_{Q(0,R)}(B_R) - d_{Q(0,R)}(A))R^d \\ &\quad + \int_{Q(0,R)} (A(x) - A_i(x)) \, dx \\ &\geq (\mathbf{m}_{Q(0,i)}(P) - 2^{-i})i^d - \frac{\varepsilon R^d}{2} + (\mathbf{m}_{\mathbb{R}^d}(P) - d_{Q(0,R)}(A))R^d \\ &\quad + \int_{Q(0,R)} (\tilde{A}(x) - A_i(x)) \, dx. \end{aligned}$$

Since  $\text{vol}(A'_i) \leq \mathbf{m}_{Q(0,i)}(P) i^d$  for all  $i \geq R$  and  $\int_{Q(0,R)} (\tilde{A}(x) - A_i(x)) \, dx > -\varepsilon$  for all sufficiently large  $i$ , we conclude that for large enough  $i$ , we have

$$d_{Q(0,R)}(A) > \mathbf{m}_{\mathbb{R}^d}(P) - \frac{i^d}{2^i R^d} - \frac{\varepsilon}{2} - \frac{\varepsilon}{R^d} > \mathbf{m}_{\mathbb{R}^d}(P) - \varepsilon,$$

proving inequality (9).

Finally, since  $I_P(A) = 0$ , it follows from Lemma 2 that we can remove a zero-measure subset of  $A$  in order to remove all copies of  $P$  without changing its density. The theorem follows.  $\square$

### 3. Configurations on the sphere

In this section, we turn to the question of whether the methods and results shown in the Euclidean space setting can also be made to work in the spherical setting.

We shall fix an integer  $d \geq 2$  throughout this section and work on the  $d$ -dimensional unit sphere  $\mathbb{S}^d := \{x \in \mathbb{R}^{d+1} : \|x\| = 1\}$ . We denote the uniform probability measure on  $\mathbb{S}^d$  by  $\sigma^{(d)} = \sigma$ , and the normalized Haar measure on  $O(\mathbb{R}^{d+1})$  by  $\mu_{d+1} = \mu$ . These two measures are related as follows: if



$X \subseteq \mathbb{S}^d$  is a measurable set and  $x \in \mathbb{S}^d$ , then

$$\sigma(X) = \mu(\{T \in O(\mathbb{R}^{d+1}) : Tx \in X\}).$$

The analogue of the axis-parallel cube in the spherical setting will be the *spherical cap*: given  $x \in \mathbb{S}^d$  and  $\rho > 0$ , we denote<sup>8</sup>

$$\text{Cap}(x, \rho) := \{y \in \mathbb{S}^d : \|x - y\|_{\mathbb{R}^{d+1}} \leq \rho\}.$$

We say  $\text{Cap}(x, \rho)$  is the *spherical cap with center  $x$  and radius  $\rho$* . Since its measure  $\sigma(\text{Cap}(x, \rho))$  does not depend on the center point  $x$ , we shall denote this value simply by  $\sigma(\text{Cap}_\rho)$ . For a given (measurable) set  $A \subseteq \mathbb{S}^d$ , we then write

$$d_{\text{Cap}(x, \rho)}(A) := \frac{\sigma(A \cap \text{Cap}(x, \rho))}{\sigma(\text{Cap}_\rho)}$$

for the density of  $A$  inside this cap.

We define a (spherical) configuration on  $\mathbb{S}^d$  as a finite subset of  $\mathbb{R}^{d+1}$  which is congruent to a set on  $\mathbb{S}^d$ ; it is convenient to allow for configurations that are not necessarily on the sphere in order to consider dilations. Note that, if  $P, Q \subset \mathbb{S}^d$  are two configurations which *are* on the sphere, then  $P \simeq Q$  if and only if there is a transformation  $T \in O(\mathbb{R}^{d+1})$  for which  $P = T \cdot Q$  (translations are no longer necessary in this case).

A spherical configuration  $P$  on  $\mathbb{S}^d$  is said to be *admissible* if it has at most  $d$  points and if it is congruent to a set  $P' \subset \mathbb{S}^d$  which is linearly independent.<sup>9</sup> As before, we shall say that some set  $A \subseteq \mathbb{S}^d$  *avoids*  $P$  if there is no subset of  $A$  which is congruent to  $P$ .

The natural analogues of the independence density in the spherical setting can now be given. For  $n \geq 1$  configurations  $P_1, \dots, P_n$  on  $\mathbb{S}^d$ , we define the quantities

$$\begin{aligned} \mathbf{m}_{\mathbb{S}^d}(P_1, \dots, P_n) &:= \sup \{ \sigma(A) : A \subset \mathbb{S}^d \text{ avoids } P_i, 1 \leq i \leq n \} \text{ and} \\ \mathbf{m}_{\text{Cap}(x, \rho)}(P_1, \dots, P_n) &:= \sup \{ d_{\text{Cap}(x, \rho)}(A) : A \subset \text{Cap}(x, \rho) \text{ avoids } P_i, 1 \leq i \leq n \}. \end{aligned}$$

Whenever convenient, we will state and prove results in the case of only one forbidden configuration, as the more general case of multiple forbidden configurations follows from the same arguments with only trivial modifications (but heavier notation).

The first issue we encounter in the spherical setting is that it is not compatible with dilations: given a set of points  $P \subset \mathbb{S}^d$  and some dilation parameter  $t > 0$ , it is usually *not* true that there exists a set  $Q \subset \mathbb{S}^d$  congruent to  $tP$ . However, there is a large class of configurations (including the ones we call admissible) for which this *is* true whenever  $0 < t \leq 1$ ; we shall say that they are *contractible*.

It is easy to show that any configuration  $P \subset \mathbb{S}^d$  which is contained in a  $d$ -dimensional affine hyperplane (e.g., any configuration with at most  $d + 1$  points) is contractible. Indeed, let  $0 < t \leq 1$  and suppose  $P \subset \mathbb{S}^d \cap (w + U)$ , where  $U \subset \mathbb{R}^{d+1}$  is a  $d$ -dimensional subspace and  $w$  is orthogonal to  $U$ . Then,  $w$  is orthogonal to  $v - w$  for every  $v \in P$ , and one readily checks that<sup>10</sup>  $sw + tP \subset \mathbb{S}^d$  for  $s = (t^2 + (1 - t^2)\|w\|^{-2})^{1/2} - t$ .

Even when the configuration we are considering is contractible, however, there is no easy relationship between the independence densities of its distinct dilates. We will then start with the following reassuring

<sup>8</sup>It is more customary to define the spherical cap using angular distance instead of Euclidean distance as we use. There is no meaningful (qualitative) difference between these two choices, but the use of the Euclidean distance will be more convenient for us.

<sup>9</sup>Note that this definition is different from the one in the Euclidean setting, where we required the points to be affinely independent instead of linearly independent. The reason behind this difference is that the Euclidean space is translation-invariant while the sphere is not, so affine properties on  $\mathbb{R}^d$  should translate to linear properties on  $\mathbb{S}^d$ .

<sup>10</sup>This is true if  $w \neq 0$ , by two applications of Pythagoras' Theorem. If  $w = 0$ , then one has instead that  $(1 - t^2)^{1/2}u + tP \subset \mathbb{S}^d$  for a unit vector  $u \in \mathbb{R}^{d+1}$  orthogonal to the subspace  $U$ .

lemma, which in a sense assures us the results we will eventually obtain are not true for only trivial reasons.

**Lemma 10.** *For any fixed contractible configuration  $P \subset \mathbb{S}^d$ , we have that*

$$\inf_{0 < t \leq 1} \mathbf{m}_{\mathbb{S}^d}(tP) > 0 \quad \text{and} \quad \sup_{0 < t \leq 1} \mathbf{m}_{\mathbb{S}^d}(tP) < 1.$$

*Proof.* For the first inequality, we note that spherical caps are exactly the closed balls of the separable metric space  $\mathbb{S}^d$  endowed with the Euclidean distance. This allows us to use the Vitali Covering Lemma; see, for instance, [14, Theorem 2.1]. For any given  $0 < t \leq 1$ , start with the trivial cover  $\mathbb{S}^d = \bigcup_{x \in \mathbb{S}^d} \text{Cap}(x, \text{diam } tP)$  and apply the Vitali Covering Lemma to obtain a (necessarily finite) set of center points  $\{x_1, \dots, x_N\} \subset \mathbb{S}^d$  such that  $\text{Cap}(x_i, \text{diam } tP) \cap \text{Cap}(x_j, \text{diam } tP) = \emptyset$  for  $i \neq j$  and

$$\mathbb{S}^d = \bigcup_{i=1}^N \text{Cap}(x_i, 5\text{diam } tP).$$

Since the caps  $\text{Cap}(x_i, \text{diam } tP)$  are pairwise disjoint, it is easy to see that the set

$$A_t := \bigcup_{i=1}^N \text{Cap}\left(x_i, \frac{\text{diam } tP}{4}\right)$$

does not contain any copy of  $tP$ . Finally, as the inequality  $\sigma(\text{Cap}_\rho) \geq c_d \sigma(\text{Cap}_{20\rho})$  holds for some constant  $c_d > 0$  and all  $0 < \rho \leq 2$ , denoting  $\rho(t) := (\text{diam } tP)/4$ , we have that

$$\sigma(A_t) = \sum_{i=1}^N \sigma(\text{Cap}_{\rho(t)}) \geq c_d \sum_{i=1}^N \sigma(\text{Cap}_{20\rho(t)}) \geq c_d \sigma\left(\bigcup_{i=1}^N \text{Cap}(x_i, 20\rho(t))\right) = c_d,$$

and thus,  $\mathbf{m}_{\mathbb{S}^d}(tP) \geq \sigma(A_t) \geq c_d$  for all  $0 < t \leq 1$ .

For the second inequality, suppose  $A \subseteq \mathbb{S}^d$  avoids  $P = \{v_1, \dots, v_k\}$ . Then,

$$\sum_{i=1}^k A(Rv_i) = |A \cap RP| \leq k - 1 \quad \text{for all } R \in \text{O}(\mathbb{R}^{d+1}).$$

Integrating over  $\text{O}(\mathbb{R}^{d+1})$ , we obtain

$$k \sigma(A) = \int_{\text{O}(\mathbb{R}^{d+1})} \left( \sum_{i=1}^k A(Rv_i) \right) d\mu(R) \leq k - 1,$$

implying that  $\sigma(A) \leq 1 - 1/k$ . Thus,  $\sup_{0 < t \leq 1} \mathbf{m}_{\mathbb{S}^d}(tP) \leq 1 - 1/|P|$ . □

Given some configuration  $P = \{v_1, v_2, \dots, v_k\} \subset \mathbb{S}^d$ , we define the *counting function*  $I_P$  which acts on a bounded measurable function  $f : \mathbb{S}^d \rightarrow \mathbb{R}$  by

$$I_P(f) := \int_{\text{O}(\mathbb{R}^{d+1})} f(Rv_1) f(Rv_2) \cdots f(Rv_k) d\mu(R).$$

In the case where  $f$  is the indicator function of a set  $A \subseteq \mathbb{S}^d$ , we note that

$$I_P(A) = \mathbb{P}_{R \in \text{O}(\mathbb{R}^{d+1})} (Rv_1, Rv_2, \dots, Rv_k \in A).$$

If the spherical configuration  $P$  is *not* a subset of the sphere, we define the function  $I_P$  as being equal to  $I_Q$  for any  $Q \simeq P$  which is contained in  $\mathbb{S}^d$ .

As in the Euclidean setting, one can show there is no meaningful difference between requiring that a measurable set  $A \subseteq \mathbb{S}^d$  avoids some configuration  $P$  or that it only satisfies  $I_P(A) = 0$ . This is proven in the next lemma.

**Lemma 11** (Zero-measure removal). *Suppose  $P \subset \mathbb{S}^d$  is a finite configuration and  $A \subseteq \mathbb{S}^d$  is measurable. If  $I_P(A) = 0$ , then we can remove a zero-measure subset of  $A$  in order to remove all copies of  $P$ .*

*Proof.* It will be more convenient to change spaces and work on the orthogonal group  $O(\mathbb{R}^{d+1})$  rather than on the sphere  $\mathbb{S}^d$ . For  $\delta > 0$  and  $R \in O(\mathbb{R}^{d+1})$ , denote by

$$\mathcal{B}(R, \delta) := \{T \in O(\mathbb{R}^{d+1}) : \|T - R\| \leq \delta\}$$

the ball of radius  $\delta$  in spectral norm centered on  $R$  and let  $I$  denote the identity transformation. We will first show that

$$\lim_{\delta \rightarrow 0} \left| \frac{1}{\mu(\mathcal{B}(I, \delta))} \int_{\mathcal{B}(I, \delta)} A(Tx) d\mu(T) - A(x) \right| = 0 \text{ for almost every } x \in \mathbb{S}^d. \tag{10}$$

Let  $e \in \mathbb{S}^d$  be an arbitrary point and define on  $O(\mathbb{R}^{d+1})$  the (measurable) set  $E := \{R \in O(\mathbb{R}^{d+1}) : Re \in A\}$ . By the Lebesgue Density Theorem on  $O(\mathbb{R}^{d+1})$ , we have that

$$\lim_{\delta \rightarrow 0} \left| \frac{1}{\mu(\mathcal{B}(R, \delta))} \int_{\mathcal{B}(R, \delta)} E(T) d\mu(T) - E(R) \right| = 0 \text{ for } \mu\text{-a.e. } R \in O(\mathbb{R}^{d+1}).$$

But this means exactly that the measure of the set

$$F := \left\{ R \in O(\mathbb{R}^{d+1}) : \lim_{\delta \rightarrow 0} \left| \frac{1}{\mu(\mathcal{B}(I, \delta))} \int_{\mathcal{B}(I, \delta)} A(TRe) d\mu(T) - A(Re) \right| \neq 0 \right\}$$

of nondensity points is zero. It is clear from the definition of  $F$  that it is invariant under the right-action of  $\text{Stab}^{O(\mathbb{R}^{d+1})}(e)$ ; this implies  $\sigma(\{Re : R \in F\}) = \mu(F) = 0$ , proving (10).

Now we remove from  $A$  all points  $x$  for which identity (10) does not hold, thus obtaining a subset  $B \subseteq A$  with  $\sigma(A \setminus B) = 0$  and

$$\lim_{\delta \rightarrow 0} \frac{1}{\mu(\mathcal{B}(I, \delta))} \int_{\mathcal{B}(I, \delta)} B(Tx) d\mu(T) = 1 \quad \text{for all } x \in B.$$

We will show that no copy of  $P$  remains on this restricted set  $B$ , which will finish the proof of the lemma.

Suppose, for contradiction, that  $B$  contains a copy  $\{u_1, \dots, u_k\}$  of  $P$ . Then, there exists  $\delta > 0$  for which

$$\frac{1}{\mu(\mathcal{B}(I, \delta))} \int_{\mathcal{B}(I, \delta)} B(Tu_i) d\mu(T) \geq 1 - \frac{1}{2k} \quad \text{for all } 1 \leq i \leq k,$$

which means that  $\mathbb{P}_{T \in \mathcal{B}(I, \delta)}(Tu_i \notin B) \leq 1/2k$  for each  $1 \leq i \leq k$ . Thus,

$$\begin{aligned} I_P(B) &= \mathbb{P}_{T \in O(\mathbb{R}^{d+1})}(Tu_1, \dots, Tu_k \in B) \\ &\geq \mu(\mathcal{B}(I, \delta)) \cdot \mathbb{P}_{T \in \mathcal{B}(I, \delta)}(Tu_1, \dots, Tu_k \in B) \\ &\geq \mu(\mathcal{B}(I, \delta)) \left( 1 - \sum_{i=1}^k \mathbb{P}_{T \in \mathcal{B}(I, \delta)}(Tu_i \notin B) \right) \\ &\geq \frac{\mu(\mathcal{B}(I, \delta))}{2} > 0, \end{aligned}$$

contradicting our assumption that  $I_P(A) = 0$ . □

### 3.1. Harmonic analysis on the sphere and the Counting Lemma

The next thing we need is an analogue of the Counting Lemma in the spherical setting, saying we do not significantly change the count of configurations in a given set  $A \subseteq \mathbb{S}^d$  by blurring this set a little. As in the Euclidean setting, we will use Fourier-analytic methods to prove such a result; we now give a quick overview of the definitions and results on harmonic analysis we will need for our arguments.

Given an integer  $n \geq 0$ , we write  $\mathcal{H}_n^{d+1}$  for the space of real harmonic polynomials, homogeneous of degree  $n$ , on  $\mathbb{R}^{d+1}$ . That is,

$$\mathcal{H}_n^{d+1} = \left\{ f \in \mathbb{R}[x_1, \dots, x_{d+1}] : f \text{ homogeneous, } \deg f = n, \sum_{i=1}^{d+1} \frac{\partial^2}{\partial x_i^2} f = 0 \right\}.$$

The restriction of the elements of  $\mathcal{H}_n^{d+1}$  to  $\mathbb{S}^d$  are called *spherical harmonics of degree  $n$*  on  $\mathbb{S}^d$ . If  $Y \in \mathcal{H}_n^{d+1}$ , note that  $Y(x) = \|x\|^n Y(x')$  where  $x = \|x\|x'$  and  $x' \in \mathbb{S}^d$ ; we can then identify  $\mathcal{H}_n^{d+1}$  with the space of spherical harmonics of degree  $n$ , which by a slight (and common) abuse of notation, we also denote  $\mathcal{H}_n^{d+1}$ .

Harmonic polynomials of different degrees are orthogonal with respect to the standard inner product  $\langle f, g \rangle_{\mathbb{S}^d} := \int_{\mathbb{S}^d} f(x)g(x) d\sigma(x)$ . Moreover, it is a well-known fact (see e.g., [5, Chapter 1.1]) that the family of spherical harmonics is dense in  $L^2(\mathbb{S}^d)$ , and so

$$L^2(\mathbb{S}^d) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n^{d+1}.$$

Denoting by  $\text{proj}_n : L^2(\mathbb{S}^d) \rightarrow \mathcal{H}_n^{d+1}$  the orthogonal projection onto  $\mathcal{H}_n^{d+1}$ , what this means is that  $f = \sum_{n=0}^{\infty} \text{proj}_n f$  for all  $f \in L^2(\mathbb{S}^d)$  (with equality in the  $L^2$  sense). By orthogonality, we obtain *Parseval's identity*:

$$\|f\|_2^2 = \sum_{n=0}^{\infty} \|\text{proj}_n f\|_2^2.$$

There is a family  $(P_n^d)_{n \geq 0}$  of polynomials on  $[-1, 1]$ , usually called *ultraspherical* or *Gegenbauer* polynomials, which is associated to this decomposition. We use the convention that  $\deg P_n^d = n$  and  $P_n^d(1) = 1$ . These polynomials can be defined via the *addition formula*

$$P_n^d(x \cdot y) = \frac{1}{\dim \mathcal{H}_n^{d+1}} \sum_{i=1}^{\dim \mathcal{H}_n^{d+1}} Y_i(x)Y_i(y) \quad \text{for all } x, y \in \mathbb{S}^d, \tag{11}$$

where  $\{Y_i : 1 \leq i \leq \dim \mathcal{H}_n^{d+1}\}$  is an (arbitrary) orthonormal basis of  $\mathcal{H}_n^{d+1}$ . We refer the reader to Chapter 1.2 of Dai and Xu's book [5] for the proof that this formula is independent of the choice of basis, and that it indeed defines a polynomial on  $[-1, 1]$ .

The next theorem collects several properties of the Gegenbauer polynomials which will be useful for us.

**Theorem 8.** *For all integers  $d \geq 2$  and  $n \geq 0$ , the following hold:*

- (i)  $P_n^d(t) \in [-1, 1]$  for all  $t \in [-1, 1]$ .
- (ii) The projection operator  $\text{proj}_n : L^2(\mathbb{S}^d) \rightarrow \mathcal{H}_n^{d+1}$  is given by

$$\text{proj}_n f(x) = \dim \mathcal{H}_n^{d+1} \int_{\mathbb{S}^d} P_n^d(x \cdot y) f(y) d\sigma(y). \tag{12}$$

(iii) For each fixed  $y, z \in \mathbb{S}^d$ , we have

$$\int_{\mathbb{S}^d} P_n^d(x \cdot y) P_n^d(x \cdot z) d\sigma(x) = \frac{1}{\dim \mathcal{H}_n^{d+1}} P_n^d(y \cdot z). \tag{13}$$

(iv) For any fixed  $\gamma > 0$ ,  $\max_{t \in [-1+\gamma, 1-\gamma]} P_n^d(t)$  tends to zero as  $n \rightarrow \infty$ .

*Proof.* The first three items follow easily from the addition formula (11). Indeed, fix some orthonormal basis  $\{Y_i : 1 \leq i \leq \dim \mathcal{H}_n^{d+1}\}$  of  $\mathcal{H}_n^{d+1}$ . Then,

$$\begin{aligned} |P_n^d(x \cdot y)| &= \left| \int_{O(d+1)} P_n^d(Rx \cdot Ry) d\mu(R) \right| \\ &= \frac{1}{\dim \mathcal{H}_n^{d+1}} \left| \int_{O(d+1)} \sum_{i=1}^{\dim \mathcal{H}_n^{d+1}} Y_i(Rx) Y_i(Ry) d\mu(R) \right|, \end{aligned}$$

which by the triangle inequality followed by Cauchy-Schwarz is at most

$$\begin{aligned} \frac{1}{\dim \mathcal{H}_n^{d+1}} \sum_{i=1}^{\dim \mathcal{H}_n^{d+1}} \left( \int_{O(d+1)} Y_i(Rx)^2 d\mu(R) \right)^{1/2} \left( \int_{O(d+1)} Y_i(Ry)^2 d\mu(R) \right)^{1/2} \\ = \frac{1}{\dim \mathcal{H}_n^{d+1}} \sum_{i=1}^{\dim \mathcal{H}_n^{d+1}} \left( \int_{\mathbb{S}^d} Y_i(z)^2 d\sigma(z) \right) = 1, \end{aligned}$$

proving (i). Item (ii) follows from the chain of equalities

$$\begin{aligned} \text{proj}_n f(x) &= \sum_{i=1}^{\dim \mathcal{H}_n^{d+1}} \left( \int_{\mathbb{S}^d} f(y) Y_i(y) d\sigma(y) \right) Y_i(x) \\ &= \int_{\mathbb{S}^d} \left( \sum_{i=1}^{\dim \mathcal{H}_n^{d+1}} Y_i(x) Y_i(y) \right) f(y) d\sigma(y) \\ &= \dim \mathcal{H}_n^{d+1} \int_{\mathbb{S}^d} P_n^d(x \cdot y) f(y) d\sigma(y). \end{aligned}$$

To prove item (iii), note that

$$\begin{aligned} \int_{\mathbb{S}^d} P_n^d(x \cdot y) P_n^d(x \cdot z) d\sigma(x) \\ &= \frac{1}{(\dim \mathcal{H}_n^{d+1})^2} \int_{\mathbb{S}^d} \sum_{i,j=1}^{\dim \mathcal{H}_n^{d+1}} Y_i(x) Y_i(y) Y_j(x) Y_j(z) d\sigma(x) \\ &= \frac{1}{(\dim \mathcal{H}_n^{d+1})^2} \sum_{i,j=1}^{\dim \mathcal{H}_n^{d+1}} \left( \int_{\mathbb{S}^d} Y_i(x) Y_j(x) d\sigma(x) \right) Y_i(y) Y_j(z) \\ &= \frac{1}{(\dim \mathcal{H}_n^{d+1})^2} \sum_{i=1}^{\dim \mathcal{H}_n^{d+1}} Y_i(y) Y_i(z), \end{aligned}$$

which equals the right-hand side of (13) by definition.

Finally, the last item immediately follows from the more precise asymptotic bound given in [21, Theorem 8.21.6]. □

We will follow Dunkl [8] in defining both the convolution operation on the sphere and the spherical analogue of Fourier coefficients. For this, we will need to break the symmetry of the sphere a little and distinguish an (arbitrary) point  $e$  on  $\mathbb{S}^d$ ; we think of this point as being the north pole. Write  $\mathcal{M}(\mathbb{S}^d; e)$  for the space of Borel regular zonal measures on  $\mathbb{S}^d$  with pole at  $e$ , that is, those measures which are invariant under the action of  $\text{Stab}^{O(\mathbb{R}^{d+1})}(e)$ . We will refer to the elements of  $\mathcal{M}(\mathbb{S}^d; e)$  simply as zonal measures.

Given a function  $f \in L^2(\mathbb{S}^d)$  and a zonal measure  $\nu \in \mathcal{M}(\mathbb{S}^d; e)$ , we define their convolution  $f * \nu$  by

$$f * \nu(x) := \int_{\mathbb{S}^d} f(T_x y) d\nu(y) \quad \text{for all } x \in \mathbb{S}^d,$$

where  $T_x \in O(\mathbb{R}^{d+1})$  is an arbitrary element satisfying  $T_x e = x$ . It is easy to see that this operation is well-defined, independently of the choice of  $T_x$ : if  $S_x e = T_x e = x$ , then  $S_x^{-1} T_x \in \text{Stab}(e)$  and so  $\nu(S_x^{-1} A) = \nu((S_x^{-1} T_x) T_x^{-1} A) = \nu(T_x^{-1} A)$ . The value  $f * \nu(x)$  can be thought of as the average of  $f$  according to a measure which acts with respect to  $x$  as  $\nu$  acts with respect to the north pole  $e$ .

For an integer  $n \geq 0$  and a zonal measure  $\nu \in \mathcal{M}(\mathbb{S}^d; e)$ , we define its  $n$ -th Fourier coefficient  $\widehat{\nu}_n$  by

$$\widehat{\nu}_n = \int_{\mathbb{S}^d} P_n^d(e \cdot y) d\nu(y).$$

The main property we will need of Fourier coefficients is the following result, which is stated in Dunkl’s paper [8] and can be proven using a straightforward modification of the methods exposed in Chapter 2 of Dai and Xu’s book [5].

**Theorem 9.** *If  $f \in L^2(\mathbb{S}^d)$  and  $\nu \in \mathcal{M}(\mathbb{S}^d; e)$ , then  $f * \nu \in L^2(\mathbb{S}^d)$  and*

$$\text{proj}_n(f * \nu) = \widehat{\nu}_n \text{proj}_n f \quad \text{for all } n \geq 0.$$

With this, we finish our review of harmonic analysis on the sphere, so let us return to our specific problem. For a given  $\delta > 0$ , denote by  $\text{cap}_\delta$  the uniform probability measure on the spherical cap  $\text{Cap}(e, \delta)$ :

$$\text{cap}_\delta(A) = \frac{\sigma(A \cap \text{Cap}(e, \delta))}{\sigma(\text{Cap}(e, \delta))} \quad \text{for all measurable } A \subseteq \mathbb{S}^d.$$

Note that each  $\text{cap}_\delta$  is a zonal measure. One immediately checks that

$$(\widehat{\text{cap}}_\delta)_n = \frac{1}{\sigma(\text{Cap}_\delta)} \int_{\text{Cap}(e, \delta)} P_n^d(e \cdot y) d\sigma(y)$$

for all  $n \geq 0$ , and

$$f * \text{cap}_\delta(x) = \frac{1}{\sigma(\text{Cap}_\delta)} \int_{\text{Cap}(x, \delta)} f(y) d\sigma(y)$$

for all  $f \in L^2(\mathbb{S}^d)$ . In particular, if  $A \subseteq \mathbb{S}^d$  is a measurable set, then  $A * \text{cap}_\delta(x) = d_{\text{Cap}(x, \delta)}(A)$ ; this gives the ‘blurring’ of the spherical sets we shall consider.

**Lemma 12.** *For every  $d \geq 2$  and  $\gamma > 0$ , there exists a function  $c_{d,\gamma} : (0, 1] \rightarrow \mathbb{R}$  with  $\lim_{\delta \rightarrow 0^+} c_{d,\gamma}(\delta) = 0$  such that the following holds: for all  $f, g \in L^2(\mathbb{S}^d)$  and all points  $u, v \in \mathbb{S}^d$  with  $|u \cdot v| \leq 1 - \gamma$ , we have that*

$$\left| \int_{O(\mathbb{R}^{d+1})} f(Ru)(g(Rv) - g * \text{cap}_\delta(Rv)) d\mu(R) \right| \leq c_{d,\gamma}(\delta) \|f\|_2 \|g\|_2.$$

*Proof.* Denote by  $\nu_e$  the Haar measure on  $\text{Stab}(e)$ , and assume without loss of generality that  $u$  coincides with the north pole  $e$ . By symmetry, the expression we wish to bound may then be written as

$$\left| \int_{\mathbb{O}(\mathbb{R}^{d+1})} f(Re)h(Rv) d\mu(R) \right| = \left| \int_{\mathbb{O}(\mathbb{R}^{d+1})} f(Re) \left( \int_{\text{Stab}(e)} h(RSv) d\nu_e(S) \right) d\mu(R) \right|,$$

where  $h = g - g * \text{cap}_\delta$ .

Write  $t_0 := e \cdot v$ . Note that when  $S \in \text{Stab}(e)$  is distributed uniformly according to  $\nu_e$ , the point  $Sv$  is uniformly distributed on  $\mathbb{S}_{t_0}^{d-1} := \{y \in \mathbb{S}^d : e \cdot y = t_0\}$ . Denote by  $\sigma_{t_0}^{(d-1)}$  the uniform probability measure on  $\mathbb{S}_{t_0}^{d-1}$  (that is, the unique one which is invariant under the action of  $\text{Stab}(e)$ ). Making the change of variables  $y = Sv$ , we see that

$$\int_{\text{Stab}(e)} h(RSv) d\nu_e(S) = \int_{\mathbb{S}_{t_0}^{d-1}} h(Ry) d\sigma_{t_0}^{(d-1)}(y) = h * \sigma_{t_0}^{(d-1)}(Re). \tag{14}$$

The expression we wish to bound is then equal to

$$\left| \int_{\mathbb{O}(\mathbb{R}^{d+1})} f(Re) h * \sigma_{t_0}^{(d-1)}(Re) d\mu(R) \right| = \left| \int_{\mathbb{S}^d} f(x) h * \sigma_{t_0}^{(d-1)}(x) d\sigma(x) \right|.$$

Using Parseval’s Identity, we can rewrite the right-hand side of the last equality as

$$\begin{aligned} & \left| \sum_{n=0}^{\infty} \int_{\mathbb{S}^d} \text{proj}_n f(x) \text{proj}_n (h * \sigma_{t_0}^{(d-1)})(x) d\sigma(x) \right| \\ & \leq \sum_{n=0}^{\infty} \int_{\mathbb{S}^d} |\text{proj}_n f(x)| |(\widehat{\sigma}_{t_0}^{(d-1)})_n| |\text{proj}_n h(x)| d\sigma(x) \\ & \leq \sum_{n=0}^{\infty} |(\widehat{\sigma}_{t_0}^{(d-1)})_n| \|\text{proj}_n f\|_2 \|\text{proj}_n h\|_2, \end{aligned}$$

where we used Theorem 9 and then Cauchy-Schwarz. As  $h = g - g * \text{cap}_\delta$ , the expression above is equal to

$$\begin{aligned} & \sum_{n=0}^{\infty} |(\widehat{\sigma}_{t_0}^{(d-1)})_n| |1 - (\widehat{\text{cap}}_\delta)_n| \|\text{proj}_n f\|_2 \|\text{proj}_n g\|_2 \\ & = \sum_{n=0}^{\infty} |P_n^d(t_0)| \left| 1 - \frac{1}{\sigma(\text{Cap}_\delta)} \int_{\text{Cap}(e, \delta)} P_n^d(e \cdot y) d\sigma(y) \right| \|\text{proj}_n f\|_2 \|\text{proj}_n g\|_2. \end{aligned}$$

Fix some  $\varepsilon > 0$ . Since  $t_0 \in [-1 + \gamma, 1 - \gamma]$  (by hypothesis), from Theorem 8, we obtain that  $|P_n^d(t_0)| \leq \varepsilon/2$  holds for all  $n \geq N(\varepsilon, \gamma)$ , while

$$\left| 1 - \frac{1}{\sigma(\text{Cap}_\delta)} \int_{\text{Cap}(e, \delta)} P_n^d(e \cdot y) d\sigma(y) \right| \leq \max_{-1 \leq t \leq 1} |1 - P_n^d(t)| = 2$$

always holds. Moreover, since each  $P_n^d$  is a polynomial satisfying  $P_n^d(1) = 1$ , we can choose  $\delta_0 = \delta_0(\varepsilon, \gamma) > 0$  small enough so that  $|1 - P_n^d(e \cdot y)| \leq \varepsilon$  holds whenever  $n < N(\varepsilon, \gamma)$  and  $y \in \text{Cap}(e, \delta_0)$ . This implies that the last sum is at most

$$\sum_{n=0}^{\infty} \varepsilon \|\text{proj}_n f\|_2 \|\text{proj}_n g\|_2 \leq \varepsilon \|f\|_2 \|g\|_2$$

whenever  $\delta \leq \delta_0(\varepsilon, \gamma)$ , finishing the proof. □



Recall that a spherical configuration  $P$  is admissible if it has at most  $d$  points and if it is congruent to a set  $P' \subset \mathbb{S}^d$  which is linearly independent. We can now give the spherical counterpart to the Counting Lemma from the last section.

**Lemma 13** (Counting Lemma). *For every admissible configuration  $P$  on  $\mathbb{S}^d$ , there exists a function  $\eta_P : (0, 1] \rightarrow (0, 1]$  with  $\lim_{\delta \rightarrow 0^+} \eta_P(\delta) = 0$  such that the following holds for all measurable sets  $A \subseteq \mathbb{S}^d$ :*

$$|I_P(A) - I_P(A * \text{cap}_\delta)| \leq \eta_P(\delta) \quad \text{for all } \delta \in (0, 1].$$

Moreover, this upper-bound function  $\eta_P$  can be made to hold uniformly over all configurations  $P'$  inside a neighborhood of  $P$ .

*Proof.* Up to congruence, we may assume  $P \subset \mathbb{S}^d$ . Similarly to what we did in the Euclidean setting, we will first obtain a uniform upper bound for

$$\left| \int_{O(\mathbb{R}^{d+1})} f_1(Tv_1) \cdots f_{k-1}(Tv_{k-1})(f_k(Tv_k) - f_k * \text{cap}_\delta(Tv_k)) d\mu(T) \right|,$$

valid whenever  $0 \leq f_1, \dots, f_k \leq 1$  are measurable functions and  $(v_1, v_2, \dots, v_k)$  is a permutation of the points of  $P$ .

Denote by  $G := \text{Stab}^{O(\mathbb{R}^{d+1})}(v_1, \dots, v_{k-2})$  the stabilizer of the first  $k - 2$  points of  $P$ , and by  $H := \text{Stab}^{O(\mathbb{R}^{d+1})}(v_1, \dots, v_{k-2}, v_{k-1}) = \text{Stab}^G(v_{k-1})$  the stabilizer of the first  $k - 1$  points of  $P$ . We can then bound the expression above by

$$\int_{O(\mathbb{R}^{d+1})} \left| \int_G f_{k-1}(TSv_{k-1})(f_k(TSv_k) - f_k * \text{cap}_\delta(TSv_k)) d\mu_G(S) \right| d\mu(T), \tag{15}$$

where  $\mu_G$  denotes the normalized Haar measure on  $G$ .

Denote  $\ell := d - k + 2 \geq 2$ . Since  $P$  is nondegenerate, we see that  $G \simeq O(\mathbb{R}^{\ell+1})$  and that both  $Gv_{k-1}$  and  $Gv_k$  are spheres of dimension  $\ell$ . Morally, we should then be able to apply the last lemma (with  $d = \ell$ ,  $f = f_{k-1}(T \cdot)$  and  $g = f_k(T \cdot)$ ) and easily conclude. However, the convolution in expression (15) above happens in  $\mathbb{S}^d$ , while that on the last lemma would happen in  $\mathbb{S}^\ell$ ; in particular, if  $k \geq 3$  so that  $\ell < d$ , all of the mass on the average defined by the convolution in (15) lies *outside* of the  $\ell$ -dimensional sphere  $Gv_k$ , so this argument cannot work. We will have to work harder to conclude.

Note that since  $Gv_k$  is an  $\ell$ -dimensional sphere while  $Hv_k$  is an  $(\ell - 1)$ -dimensional sphere (which happens because  $P$  is nondegenerate), it follows that there is a point  $\xi \in Gv_k$  which is fixed by  $H$ ; this point will work as the north pole of  $Gv_k$ .

It will be more convenient to work on the canonical unit sphere  $\mathbb{S}^\ell$  instead of the  $\ell$ -dimensional sphere  $Gv_k \subset \mathbb{S}^d$ . We shall then restrict ourselves to the  $(\ell + 1)$ -dimensional affine hyperplane  $\mathcal{H}$  determined by  $\mathcal{H} \cap \mathbb{S}^d = Gv_k$ , and place coordinates on it to identify  $\mathcal{H}$  with  $\mathbb{R}^{\ell+1}$  and  $Gv_k$  with  $\mathbb{S}^\ell$ , noting that  $G$  then acts as  $O(\mathbb{R}^{\ell+1})$ . More formally, let  $r > 0$  be the radius of  $Gv_k$  in  $\mathbb{R}^{d+1}$  so that  $Gv_k$  is isometric to  $r\mathbb{S}^\ell$ ; take such an isometry  $\psi : Gv_k \rightarrow r\mathbb{S}^\ell$ , and define  $e \in \mathbb{S}^\ell$  by  $e := \psi(\xi)/r$ . Now, we construct a map  $\phi : G \rightarrow O(\mathbb{R}^{\ell+1})$  defined by

$$\phi(S)\psi(x) = \psi(Sx) \quad \text{for all } x \in Gv_k$$

for each  $S \in G$ . It is easy to check that this map is well-defined and gives an isomorphism between  $G$  and  $O(\mathbb{R}^{\ell+1})$ , satisfying  $\phi(H) = \text{Stab}^{O(\mathbb{R}^{\ell+1})}(e)$ .

For each fixed  $T \in O(\mathbb{R}^{d+1})$ , define the functions  $g_T, h_T : \mathbb{S}^\ell \rightarrow [-1, 1]$  by

$$\begin{aligned} g_T(Re) &:= f_{k-1}(T\phi^{-1}(R)v_{k-1}) \quad \text{and} \\ h_T(Re) &:= f_k(T\phi^{-1}(R)\xi) - f_k * \text{cap}_\delta(T\phi^{-1}(R)\xi), \end{aligned}$$

for all  $R \in O(\mathbb{R}^{\ell+1})$ . These functions are indeed well-defined on  $\mathbb{S}^\ell$  since  $\text{Stab}^G(v_{k-1}) = \text{Stab}^G(\xi) = \phi^{-1}(\text{Stab}^{O(\mathbb{R}^{\ell+1})}(e))$ . Note that  $h_T$  can also be written as a function of  $x \in \mathbb{S}^\ell$  by making use of the isometry  $\psi^{-1} : r\mathbb{S}^\ell \rightarrow Gv_k$ :

$$h_T(x) = f_k(T\psi^{-1}(rx)) - f_k * \text{cap}_\delta(T\psi^{-1}(rx)).$$

Denote by  $u := \psi(v_k)/r$  the point in  $\mathbb{S}^\ell$  corresponding to  $v_k$ . Making the change of variables  $R = \phi(S)$ , we obtain

$$\begin{aligned} & \int_G f_{k-1}(TSv_{k-1})(f_k(TSv_k) - f_k * \text{cap}_\delta(TSv_k)) \, d\mu_G(S) \\ &= \int_{O(\mathbb{R}^{\ell+1})} g_T(Re) h_T(Ru) \, d\mu_{\ell+1}(R) \\ &= \int_{O(\mathbb{R}^{\ell+1})} g_T(Re) \left( \int_{\text{Stab}(e)} h_T(RSu) \, dv_e(S) \right) d\mu_{\ell+1}(R), \end{aligned}$$

where we write  $\text{Stab}(e)$  for  $\text{Stab}^{O(\mathbb{R}^{\ell+1})}(e)$  and  $v_e$  for its Haar measure. Working as we did to obtain equation (14), we see that the expression in parentheses is equal to  $h_T * \sigma_{e \cdot u}^{(\ell-1)}(Re)$ , where  $\sigma_{e \cdot u}^{(\ell-1)}$  is the uniform probability measure on the  $(\ell - 1)$ -sphere  $\text{Stab}(e)u = \{y \in \mathbb{S}^\ell : e \cdot y = e \cdot u\}$  (and the convolution now takes place in  $\mathbb{S}^\ell$  with  $e$  as the north pole). Making the change of variables  $x = Re$ , we then see that the expression above is equal to

$$\int_{O(\mathbb{R}^{\ell+1})} g_T(Re) h_T * \sigma_{e \cdot u}^{(\ell-1)}(Re) \, d\mu_{\ell+1}(R) = \int_{\mathbb{S}^\ell} g_T(x) h_T * \sigma_{e \cdot u}^{(\ell-1)}(x) \, d\sigma^{(\ell)}(x).$$

We conclude that the expression (15) we wish to bound is equal to

$$\begin{aligned} & \int_{O(\mathbb{R}^{d+1})} \left| \int_{\mathbb{S}^\ell} g_T(x) h_T * \sigma_{e \cdot u}^{(\ell-1)}(x) \, d\sigma^{(\ell)}(x) \right| d\mu_{d+1}(T) \\ & \leq \int_{O(\mathbb{R}^{d+1})} \|h_T * \sigma_{e \cdot u}^{(\ell-1)}\|_2 \, d\mu_{d+1}(T) \\ & \leq \left( \int_{O(\mathbb{R}^{d+1})} \|h_T * \sigma_{e \cdot u}^{(\ell-1)}\|_2^2 \, d\mu_{d+1}(T) \right)^{1/2}, \end{aligned}$$

where we applied Cauchy-Schwarz twice.

Let us now compute  $e \cdot u$ , which will be necessary for bounding  $\|h_T * \sigma_{e \cdot u}^{(\ell-1)}\|_2^2$ . From the identity

$$\|re - ru\|_{\mathbb{R}^{\ell+1}}^2 = \|\psi^{-1}(re) - \psi^{-1}(ru)\|_{\mathbb{R}^{d+1}}^2 = \|\xi - v_k\|_{\mathbb{R}^{d+1}}^2,$$

we conclude that  $r^2(2 - 2e \cdot u) = 2 - 2\xi \cdot v_k$ , and so

$$e \cdot u = (\xi \cdot v_k - (1 - r^2))/r^2 \notin \{-1, 1\}$$

depends only on the ordering  $(v_1, \dots, v_k)$  of  $P$  and not on our later choices (note that this value depends continuously on the points  $v_1, \dots, v_k$  and so is bounded away from  $\{-1, 1\}$  uniformly over all configurations  $P'$  close enough to  $P$ ).

Now fix an arbitrary  $\varepsilon > 0$ . By Parseval’s Identity and Theorem 9, we have that

$$\begin{aligned} \|h_T * \sigma_{e \cdot u}^{(\ell-1)}\|_2^2 &= \sum_{n=0}^{\infty} \|\text{proj}_n(h_T * \sigma_{e \cdot u}^{(\ell-1)})\|_2^2 \\ &= \sum_{n=0}^{\infty} |(\widehat{\sigma}_{e \cdot u}^{(\ell-1)})_n|^2 \|\text{proj}_n h_T\|_2^2 \\ &= \sum_{n=0}^{\infty} P_n^\ell(e \cdot u)^2 \|\text{proj}_n h_T\|_2^2. \end{aligned}$$

Since  $e \cdot u \notin \{-1, 1\}$  is a constant depending only on  $P$ , by Theorem 8, there exists  $N = N(\varepsilon, P) \in \mathbb{N}$  such that  $|P_n^\ell(e \cdot u)| \leq \varepsilon$  for all  $n > N$  (by that same theorem, this value of  $N$  can be made robust to small perturbations of the value  $e \cdot u$ , which corresponds to small perturbations of the configuration  $P$ ). Using that  $|P_n^\ell(t)| \leq 1$  for all  $-1 \leq t \leq 1$ , we conclude that

$$\|h_T * \sigma_{e \cdot u}^{(\ell-1)}\|_2^2 \leq \sum_{n=0}^N \|\text{proj}_n h_T\|_2^2 + \sum_{n>N} \varepsilon^2 \|\text{proj}_n h_T\|_2^2.$$

The second term on the right-hand side of the inequality above is upper bounded by  $\varepsilon^2 \|h_T\|_2^2 \leq \varepsilon^2$ , so let us concentrate on the first term.

By identities (12) and (13), we have

$$\begin{aligned} \|\text{proj}_n h_T\|_2^2 &= \int_{\mathbb{S}^\ell} \left( \dim \mathcal{H}_n^{\ell+1} \int_{\mathbb{S}^\ell} h_T(y) P_n^\ell(x \cdot y) d\sigma(y) \right)^2 d\sigma(x) \\ &= (\dim \mathcal{H}_n^{\ell+1})^2 \int_{\mathbb{S}^\ell} \int_{\mathbb{S}^\ell} h_T(y) h_T(z) \left( \int_{\mathbb{S}^\ell} P_n^\ell(x \cdot y) P_n^\ell(x \cdot z) d\sigma(x) \right) d\sigma(y) d\sigma(z) \\ &= \dim \mathcal{H}_n^{\ell+1} \int_{\mathbb{S}^\ell} \int_{\mathbb{S}^\ell} h_T(y) h_T(z) P_n^\ell(y \cdot z) d\sigma(y) d\sigma(z). \end{aligned}$$

Since  $|P_n^\ell(y \cdot z)| \leq 1$  for all  $y, z \in \mathbb{S}^\ell$ , we conclude that

$$\begin{aligned} &\int_{\text{O}(d+1)} \|\text{proj}_n h_T\|_2^2 d\mu_{d+1}(T) \\ &= \dim \mathcal{H}_n^{\ell+1} \int_{\mathbb{S}^\ell} \int_{\mathbb{S}^\ell} \left( \int_{\text{O}(d+1)} h_T(y) h_T(z) d\mu_{d+1}(T) \right) P_n^\ell(y \cdot z) d\sigma(y) d\sigma(z) \\ &\leq \dim \mathcal{H}_n^{\ell+1} \int_{\mathbb{S}^\ell} \int_{\mathbb{S}^\ell} \left| \int_{\text{O}(d+1)} h_T(y) h_T(z) d\mu_{d+1}(T) \right| d\sigma(y) d\sigma(z). \end{aligned}$$

We now divide this last double integral on the sphere into two parts, depending on whether or not  $y \cdot z$  is close to the extremal points 1 or  $-1$ . Thus, for some parameter  $0 < \gamma < 1$  to be chosen later, we write the double integral as

$$\begin{aligned} &\int_{\mathbb{S}^\ell} \int_{\mathbb{S}^\ell} \left| \int_{\text{O}(d+1)} h_T(y) h_T(z) d\mu_{d+1}(T) \right| \mathbb{1}\{|y \cdot z| > 1 - \gamma\} d\sigma(y) d\sigma(z) \\ &\quad + \int_{\mathbb{S}^\ell} \int_{\mathbb{S}^\ell} \left| \int_{\text{O}(d+1)} h_T(y) h_T(z) d\mu_{d+1}(T) \right| \mathbb{1}\{|y \cdot z| \leq 1 - \gamma\} d\sigma(y) d\sigma(z). \end{aligned}$$

Since  $-1 \leq h_T \leq 1$ , the first term is at most

$$2 \int_{\mathbb{S}^\ell} \int_{\mathbb{S}^\ell} \mathbb{1}\{y \cdot z > 1 - \gamma\} d\sigma(y) d\sigma(z) = 2\sigma^{(\ell)}(\text{Cap}_{\mathbb{S}^\ell}(e, \sqrt{2\gamma})).$$

To bound the second term, note that for fixed  $y, z \in \mathbb{S}^\ell$ , we have

$$\begin{aligned} & \int_{\mathbb{O}(d+1)} h_T(y)h_T(z) d\mu_{d+1}(T) \\ &= \int_{\mathbb{O}(d+1)} (f_k(T\tilde{y}) - f_k * \text{cap}_\delta(T\tilde{y}))(f_k(T\tilde{z}) - f_k * \text{cap}_\delta(T\tilde{z})) d\mu_{d+1}(T), \end{aligned}$$

where  $\tilde{y} := \psi^{-1}(ry)$  and  $\tilde{z} := \psi^{-1}(rz)$ . Moreover, we have

$$\|ry - rz\|_{\mathbb{R}^{\ell+1}}^2 = \|\tilde{y} - \tilde{z}\|_{\mathbb{R}^{d+1}}^2 \implies \tilde{y} \cdot \tilde{z} = 1 - r^2(1 - y \cdot z);$$

thus, whenever  $|y \cdot z| \leq 1 - \gamma$ , we have  $|\tilde{y} \cdot \tilde{z}| \leq 1 - r^2\gamma$ . Using Lemma 12 (with  $f = f_k - f_k * \text{cap}_\delta$ ,  $g = f_k$  and  $\gamma$  substituted by  $r^2\gamma$ ), we conclude that the second term is bounded by  $c_{d,r^2\gamma}(\delta)$ .

Taking stock of everything, we obtain

$$\begin{aligned} & \int_{\mathbb{O}(d+1)} \|h_T * \sigma_{e \cdot u}^{(\ell-1)}\|_2^2 d\mu_{d+1}(T) \\ & \leq \varepsilon^2 + \sum_{n=0}^N \dim \mathcal{H}_n^{\ell+1} (2\sigma^{(\ell)}(\text{Cap}_{\mathbb{S}^\ell}(e, \sqrt{2\gamma})) + c_{d,r^2\gamma}(\delta)) \end{aligned}$$

for any  $0 < \gamma < 1$ . Choosing  $\gamma$  small enough depending on  $\ell, \varepsilon$  and  $N$ , and then choosing  $\delta$  small enough depending on  $d, r^2\gamma, \varepsilon$  and  $N$  (so ultimately only on  $\varepsilon$  and  $P$ ), we can bound the right-hand side above by  $4\varepsilon^2$ ; the expression (15) is then bounded by  $2\varepsilon$  in this case.

For such small values of  $\delta$ , we thus conclude from our telescoping sum trick (explained in Section 2.1) that  $|I_P(A) - I_P(A * \text{cap}_\delta)| \leq 2k\varepsilon$ , proving the desired inequality since  $\varepsilon > 0$  is arbitrary. The claim that the upper bound can be made uniform inside some neighborhood of  $P$  follows from analyzing our proof.  $\square$

We remark that the proof of the Counting Lemma given above is the only place where we explicitly make use of the assumption that a spherical configuration is admissible. This assumption, however, will get inherited by all later results which make use of the Counting Lemma in their proofs.

### 3.2. Continuity properties of the counting function

Following the same script as in the Euclidean setting, we now consider other ways in which the counting function is robust to small perturbations.

It is again easy to show, using our telescoping sum trick, that  $I_P$  is continuous in  $L^\infty(\mathbb{S}^d)$  (and even in  $L^{|P|}(\mathbb{S}^d)$ ) for all spherical configurations. When the configuration considered is admissible, we obtain also the following significantly stronger continuity property of  $I_P$  when restricting to bounded functions.

**Lemma 14** (Weak\* continuity). *If  $P$  is an admissible configuration on  $\mathbb{S}^d$ , then  $I_P$  is weak\* continuous on the unit ball of  $L^\infty(\mathbb{S}^d)$ .*

*Proof.* Denote the closed unit ball of  $L^\infty(\mathbb{S}^d)$  by  $\mathcal{B}_\infty$ , and let  $(f_i)_{i \geq 1} \subset \mathcal{B}_\infty$  be a sequence weak\* converging to  $f \in \mathcal{B}_\infty$ . It will suffice to show that  $(I_P(f_i))_{i \geq 1}$  converges to  $I_P(f)$ .

Note that, for every  $x \in \mathbb{S}^d$ ,  $\delta > 0$ , we have

$$f_i * \text{cap}_\delta(x) = \frac{1}{\sigma(\text{Cap}_\delta)} \int_{\text{Cap}(x, \delta)} f_i(y) \, d\sigma(y)$$

$$\xrightarrow{i \rightarrow \infty} \frac{1}{\sigma(\text{Cap}_\delta)} \int_{\text{Cap}(x, \delta)} f(y) \, d\sigma(y) = f * \text{cap}_\delta(x).$$

Since  $f * \text{cap}_\delta$  and each  $f_i * \text{cap}_\delta$  are Lipschitz with the same constant (depending only on  $\delta$ ) and  $\mathbb{S}^d$  is compact, this easily implies that

$$\|f_i * \text{cap}_\delta - f * \text{cap}_\delta\|_\infty \rightarrow 0 \quad \text{as } i \rightarrow \infty.$$

In particular, we conclude  $\lim_{i \rightarrow \infty} I_P(f_i * \text{cap}_\delta) = I_P(f * \text{cap}_\delta)$ .

Since  $P$  is admissible, by the spherical Counting Lemma, we have

$$|I_P(f * \text{cap}_\delta) - I_P(f)| \leq \eta_P(\delta) \quad \text{and} \quad |I_P(f_i * \text{cap}_\delta) - I_P(f_i)| \leq \eta_P(\delta) \quad \text{for all } i \geq 1.$$

Choosing  $i_0(\delta) \geq 1$  sufficiently large so that

$$|I_P(f_i * \text{cap}_\delta) - I_P(f * \text{cap}_\delta)| \leq \eta_P(\delta) \quad \text{for all } i \geq i_0(\delta),$$

we conclude that

$$|I_P(f) - I_P(f_i)| \leq |I_P(f) - I_P(f * \text{cap}_\delta)| + |I_P(f * \text{cap}_\delta) - I_P(f_i * \text{cap}_\delta)|$$

$$+ |I_P(f_i * \text{cap}_\delta) - I_P(f_i)|$$

$$\leq 3\eta_P(\delta) \quad \text{for all } i \geq i_0(\delta).$$

Since  $\delta > 0$  is arbitrary and  $\eta_P(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ , this finishes the proof. □

Given some spherical configuration  $P = \{v_1, \dots, v_k\} \subset \mathbb{R}^{d+1}$ , let us write  $\mathcal{B}(P, r) \subset (\mathbb{R}^{d+1})^k$  for the ball of radius  $r$  centered on  $P$ , where the distance from  $P$  to  $Q = \{u_1, \dots, u_k\}$  is given by

$$\|Q - P\|_\infty := \min_{\sigma \in \mathfrak{S}_k} \max_{1 \leq i \leq k} \|u_i - v_{\sigma(i)}\|.$$

If  $P$  is an admissible spherical configuration, note that all configurations inside a small enough ball centered on  $P$  will also be admissible.

We will later need an equicontinuity property for the family of counting functions  $P \mapsto I_P(A)$  over all measurable sets  $A \subseteq \mathbb{S}^d$ ; this is given in the following lemma.

**Lemma 15** (Equicontinuity). *For every admissible  $P \subset \mathbb{S}^d$  and every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that*

$$|I_Q(A) - I_P(A)| \leq \varepsilon \quad \text{for all } Q \in \mathcal{B}(P, \delta), A \subseteq \mathbb{S}^d.$$

*Proof.* We will use the fact that the function  $\eta_P$  obtained in the Counting Lemma can be made uniform inside a small ball centered on  $P$ . In other words, there is  $r > 0$  and a function  $\eta'_P : (0, 1] \rightarrow (0, 1]$  with  $\lim_{t \rightarrow 0} \eta'_P(t) = 0$  such that

$$|I_Q(A) - I_Q(A * \text{cap}_\rho)| \leq \eta'_P(\rho) \quad \text{for all } Q \in \mathcal{B}(P, r), A \subseteq \mathbb{S}^d.$$

Now, for a given  $\rho > 0$  and all  $0 < \delta < \rho$ , we see from the triangle inequality that

$$\|x - y\| \leq \delta \implies \text{Cap}(x, \rho - \delta) \subset \text{Cap}(x, \rho) \cap \text{Cap}(y, \rho),$$

and so  $\sigma(\text{Cap}(x, \rho) \setminus \text{Cap}(y, \rho)) \leq \sigma(\text{Cap}_\rho) - \sigma(\text{Cap}_{\rho-\delta})$ . This implies that, for any set  $A \subseteq \mathbb{S}^d$  and any  $x, y \in \mathbb{S}^d$  with  $\|x - y\| \leq \delta$ , we have

$$\begin{aligned} |A * \text{cap}_\rho(x) - A * \text{cap}_\rho(y)| &= \frac{|\sigma(A \cap \text{Cap}(x, \rho)) - \sigma(A \cap \text{Cap}(y, \rho))|}{\sigma(\text{Cap}_\rho)} \\ &\leq \frac{\sigma(\text{Cap}(x, \rho) \setminus \text{Cap}(y, \rho))}{\sigma(\text{Cap}_\rho)} \\ &\leq \frac{\sigma(\text{Cap}_\rho) - \sigma(\text{Cap}_{\rho-\delta})}{\sigma(\text{Cap}_\rho)}. \end{aligned}$$

By our telescoping sum trick, whenever  $\|Q - P\|_\infty \leq \delta$ , we conclude that

$$|I_Q(A * \text{cap}_\rho) - I_P(A * \text{cap}_\rho)| \leq k \frac{\sigma(\text{Cap}_\rho) - \sigma(\text{Cap}_{\rho-\delta})}{\sigma(\text{Cap}_\rho)}.$$

Take  $\rho > 0$  small enough so that  $\eta'_P(\rho) \leq \varepsilon/3$ , and for this value of  $\rho$  take  $0 < \delta < r$  small enough so that  $\sigma(\text{Cap}_{\rho-\delta}) \geq (1 - \varepsilon/3k) \sigma(\text{Cap}_\rho)$ . Then, for any  $Q \in \mathcal{B}(P, \delta)$  and any measurable set  $A \subseteq \mathbb{S}^d$ , we have

$$\begin{aligned} |I_Q(A) - I_P(A)| &\leq |I_Q(A) - I_Q(A * \text{cap}_\rho)| + |I_Q(A * \text{cap}_\rho) - I_P(A * \text{cap}_\rho)| \\ &\quad + |I_P(A * \text{cap}_\rho) - I_P(A)| \\ &\leq \eta'_P(\rho) + k \frac{\sigma(\text{Cap}_\rho) - \sigma(\text{Cap}_{\rho-\delta})}{\sigma(\text{Cap}_\rho)} + \eta'_P(\rho) \\ &\leq \frac{\varepsilon}{3} + k \frac{\varepsilon}{3k} + \frac{\varepsilon}{3} = \varepsilon, \end{aligned}$$

as wished. □

### 3.3. The spherical Supersaturation Theorem

Having proven that the counting function for admissible spherical configurations is robust to various kinds of small perturbations, we next show that it also satisfies a useful *supersaturation property*.

This is the second main technical tool we need to study the independence density in the spherical setting, and due to the fact that the unit sphere is compact, both its statement and proof are somewhat simpler than in the Euclidean space setting.

**Theorem 10** (Supersaturation Theorem). *For every admissible configuration  $P$  on  $\mathbb{S}^d$  and every  $\varepsilon > 0$ , there exists a constant  $c(\varepsilon) > 0$  such that the following holds: if  $A \subseteq \mathbb{S}^d$  satisfies  $\sigma(A) \geq \mathbf{m}_{\mathbb{S}^d}(P) + \varepsilon$ , then  $I_P(A) \geq c(\varepsilon)$ .*

*Proof.* Suppose, for contradiction, that the result is false; then, there exist some  $\varepsilon > 0$  and some sequence  $(A_i)_{i \geq 1}$  of sets, each of density at least  $\mathbf{m}_{\mathbb{S}^d}(P) + \varepsilon$ , which satisfy  $\lim_{i \rightarrow \infty} I_P(A_i) = 0$ .

Note that the unit ball  $\mathcal{B}_\infty$  of  $L^\infty(\mathbb{S}^d)$  is weak\* compact and also metrizable in this topology (see [15, Chapter 2.6]). By possibly restricting to a subsequence, we may then assume that  $(A_i)_{i \geq 1}$  converges in the weak\* topology of  $L^\infty(\mathbb{S}^d)$ ; let us denote its limit by  $A \in \mathcal{B}_\infty$ . It is clear that  $0 \leq A \leq 1$  almost everywhere, and  $\int_{\mathbb{S}^d} A(x) d\sigma(x) = \lim_{i \rightarrow \infty} \sigma(A_i) \geq \mathbf{m}_{\mathbb{S}^d}(P) + \varepsilon$ . By weak\* continuity of  $I_P$  (Lemma 14), we also have  $I_P(A) = \lim_{i \rightarrow \infty} I_P(A_i) = 0$ .

Now let  $B := \{x \in \mathbb{S}^d : A(x) \geq \varepsilon\}$ . Since

$$\varepsilon B(x) \leq A(x) < \varepsilon + B(x) \quad \text{for a.e. } x \in \mathbb{S}^d,$$

we conclude that  $I_P(B) \leq \varepsilon^{-|P|} I_P(A) = 0$  and

$$\sigma(B) > \int_{\mathbb{S}^d} A(x) d\sigma(x) - \varepsilon \geq \mathbf{m}_{\mathbb{S}^d}(P).$$

But this set  $B$  contradicts Lemma 11, finishing the proof. □

It will be useful to also introduce a spherical analogue of the zooming-out operator, which acts on measurable spherical sets and represents the points on the sphere around which the considered set has a somewhat high density. Given quantities  $\delta, \gamma > 0$ , we denote by  $\mathcal{Z}_\delta(\gamma)$  the operator which takes a measurable set  $A \subseteq \mathbb{S}^d$  to the set

$$\mathcal{Z}_\delta(\gamma)[A] := \{x \in \mathbb{S}^d : d_{\text{Cap}(x, \delta)}(A) \geq \gamma\}.$$

The most important property of the zooming-out operator is the following result.

**Corollary 2.** *For every admissible configuration  $P$  on  $\mathbb{S}^d$  and every  $\varepsilon > 0$ , there exists  $\delta_0 > 0$  such that the following holds for all  $\delta \leq \delta_0$ : if  $A \subseteq \mathbb{S}^d$  satisfies*

$$\sigma(\mathcal{Z}_\delta(\varepsilon)[A]) \geq \mathbf{m}_{\mathbb{S}^d}(P) + \varepsilon,$$

*then  $A$  contains a congruent copy of  $P$ .*

*Proof.* By the Supersaturation Theorem, we know that

$$\sigma(\mathcal{Z}_\delta(\varepsilon)[A]) \geq \mathbf{m}_{\mathbb{S}^d}(P) + \varepsilon \implies I_P(\mathcal{Z}_\delta(\varepsilon)[A]) \geq c(\varepsilon)$$

holds for all  $\delta > 0$ . By the Counting Lemma, we then have

$$\begin{aligned} I_P(A) &\geq I_P(A * \text{cap}_\delta) - \eta_P(\delta) \geq \varepsilon^{|P|} I_P(\mathcal{Z}_\delta(\varepsilon)[A]) - \eta_P(\delta) \\ &\geq \varepsilon^{|P|} c(\varepsilon) - \eta_P(\delta). \end{aligned}$$

Since  $\eta_P(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ , there is some  $\delta_0 > 0$  such that for all  $\delta \leq \delta_0$ , we can conclude  $I_P(A) > 0$ ; this implies that  $A$  contains a copy of  $P$ . □

### 3.4. From the sphere to spherical caps

We must now tackle the problem of obtaining a relationship between the independence density  $\mathbf{m}_{\mathbb{S}^d}(P)$  of a given configuration  $P \subset \mathbb{S}^d$  and its spherical cap version  $\mathbf{m}_{\text{Cap}(x, \rho)}(P)$ , as this will be needed later.

In the Euclidean setting, this was very easy to do (see Lemma 1), using the fact that we can tessellate  $\mathbb{R}^d$  with cubes  $Q(x, R)$  of any given side length  $R > 0$ . This is no longer the case in the spherical setting, as it is impossible to completely cover  $\mathbb{S}^d$  using nonoverlapping spherical caps of some given radius; in fact, this cannot be done even approximately if we require the radii of the spherical caps to be the same (as we did with the side length of the cubes in  $\mathbb{R}^d$ ).

We will then need to use a much weaker ‘almost-covering’ result, saying that we can cover almost all of the sphere by using finitely many nonoverlapping spherical caps with possibly different radii. Such a collection of disjoint spherical caps is called a *cap packing*. For technical reasons, we will also want the radii of the caps in this packing to be arbitrarily small.

**Lemma 16.** *For every  $\varepsilon > 0$ , there is a finite cap packing*

$$\mathcal{P} = \{\text{Cap}(x_i, \rho_i) : 1 \leq i \leq N\}$$

*of  $\mathbb{S}^d$  with density  $\sigma(\mathcal{P}) > 1 - \varepsilon$  and with radii  $\rho_i \leq \varepsilon$  for all  $1 \leq i \leq N$ .*



*Proof.* We will use the same notation for both a collection of caps and the set of points on  $\mathbb{S}^d$  which belong to (at least) one of these caps. The desired packing  $\mathcal{P}$  will be constructed in several steps, starting with  $\mathcal{P}_0 := \{\text{Cap}(e, \varepsilon)\}$ .

Now, suppose  $\mathcal{P}_{i-1}$  has already been constructed (and is finite) for some  $i \geq 1$  and let us construct  $\mathcal{P}_i$ . Define

$$\mathcal{C}_i := \{\text{Cap}(x, \min\{\varepsilon, \text{dist}(x, \mathcal{P}_{i-1})\}) : x \in \mathbb{S}^d \setminus \mathcal{P}_{i-1}\},$$

and note that  $\mathcal{C}_i$  is a covering of  $\mathbb{S}^d \setminus \mathcal{P}_{i-1}$  by caps of positive radii (since  $\mathcal{P}_{i-1}$  is closed on  $\mathbb{S}^d$ ). By the Vitali Covering Lemma, there is a countable subcollection

$$\mathcal{Q}_i = \bigcup_{j=1}^{\infty} \{\text{Cap}(x_j, r_j)\} \subset \mathcal{C}_i$$

of disjoint caps in  $\mathcal{C}_i$  such that  $\mathbb{S}^d \setminus \mathcal{P}_{i-1} \subseteq \bigcup_{j=1}^{\infty} \text{Cap}(x_j, 5r_j)$ . In particular,

$$1 - \sigma(\mathcal{P}_{i-1}) = \sigma(\mathbb{S}^d \setminus \mathcal{P}_{i-1}) \leq \sum_{j=1}^{\infty} \sigma(\text{Cap}(x_j, 5r_j)) \leq K_d \sigma(\mathcal{Q}_i),$$

where we denote  $K_d := \sup_{r>0} \sigma(\text{Cap}_{5r})/\sigma(\text{Cap}_r) < \infty$ . Taking  $N_i \in \mathbb{N}$  such that

$$\sum_{j=1}^{N_i} \sigma(\text{Cap}(x_j, r_j)) \geq \sigma(\mathcal{Q}_i) - \frac{1 - \sigma(\mathcal{P}_{i-1})}{2K_d},$$

we see that  $\mathcal{P}'_i := \{\text{Cap}(x_j, r_j) : 1 \leq j \leq N_i\} \subset \mathbb{S}^d \setminus \mathcal{P}_{i-1}$  satisfies

$$\sigma(\mathcal{P}'_i) \geq \frac{1 - \sigma(\mathcal{P}_{i-1})}{2K_d}.$$

Now, set  $\mathcal{P}_i := \mathcal{P}_{i-1} \cup \mathcal{P}'_i$ ; this is a finite cap packing with

$$\begin{aligned} 1 - \sigma(\mathcal{P}_i) &= 1 - \sigma(\mathcal{P}_{i-1}) - \sigma(\mathcal{P}'_i) \leq (1 - \sigma(\mathcal{P}_{i-1})) \left(1 - \frac{1}{2K_d}\right) \\ &\leq (1 - \sigma(\text{Cap}_\varepsilon)) \left(1 - \frac{1}{2K_d}\right)^i \end{aligned}$$

(where the last inequality follows by induction). Taking  $n \geq 1$  large enough so that  $(1 - \sigma(\text{Cap}_\varepsilon)) \left(1 - \frac{1}{2K_d}\right)^n < \varepsilon$ , we see that  $\mathcal{P} := \mathcal{P}_n$  satisfies all requirements. □

We can now obtain our analogue of Lemma 1, relating the two versions of independence density in the spherical setting.

**Lemma 17.** *For every  $\varepsilon > 0, \rho > 0$ , there exists  $t_0 > 0$  such that the following holds whenever  $P_1, \dots, P_n \subset \mathbb{S}^d$  have diameter at most  $t_0$ :*

$$\left| \mathbf{m}_{\text{Cap}(x, \rho)}(P_1, \dots, P_n) - \mathbf{m}_{\mathbb{S}^d}(P_1, \dots, P_n) \right| < \varepsilon.$$

*Proof.* If  $A \subset \mathbb{S}^d$  is a set which avoids  $P_1, \dots, P_n$ , then for every  $x \in \mathbb{S}^d$ , the set  $A \cap \text{Cap}(x, \rho) \subseteq \text{Cap}(x, \rho)$  also avoids  $P_1, \dots, P_n$ . Since  $\mathbb{E}_{x \in \mathbb{S}^d} [d_{\text{Cap}(x, \rho)}(A)] = \sigma(A)$ , there must exist some  $x \in \mathbb{S}^d$

such that

$$d_{\text{Cap}(x,\rho)}(A \cap \text{Cap}(x, \rho)) = d_{\text{Cap}(x,\rho)}(A) \geq \sigma(A);$$

optimizing over  $A$ , we conclude that  $\mathbf{m}_{\text{Cap}(x,\rho)}(P_1, \dots, P_n) \geq \mathbf{m}_{\mathbb{S}^d}(P_1, \dots, P_n)$ .

For the opposite direction, let  $\gamma \leq \varepsilon/4$  be small enough so that  $\sigma(\text{Cap}_{\rho+\gamma}) \leq (1 + \varepsilon/4) \sigma(\text{Cap}_\rho)$ . By Lemma 16, we know there is a cap packing

$$\mathcal{P} = \{\text{Cap}(x_i, \rho_i) : 1 \leq i \leq N\}$$

of  $\mathbb{S}^d$  with  $\sigma(\mathcal{P}) \geq 1 - \gamma$  and  $0 < \rho_1, \dots, \rho_N \leq \gamma$ . Now, let  $t_0 > 0$  be small enough so that  $\sigma(\text{Cap}_{\rho_i-2t_0}) \geq (1 - \varepsilon/4) \sigma(\text{Cap}_{\rho_i})$  for all  $1 \leq i \leq N$ ; note that  $t_0$  will ultimately depend only on  $\varepsilon$  and  $\rho$ .

Fixing any configurations  $P_1, \dots, P_n \subset \mathbb{S}^d$  of diameter at most  $t_0$ , let  $A \subset \text{Cap}(x, \rho)$  be a set which avoids all of them. We shall construct a set  $\tilde{A} \subset \mathbb{S}^d$  which also avoids  $P_1, \dots, P_n$ , and which satisfies  $\sigma(\tilde{A}) > d_{\text{Cap}(x,\rho)}(A) - \varepsilon$ ; this will finish the proof.

For each  $1 \leq i \leq N$ , denote  $\tilde{\rho}_i := \rho_i - 2t_0 < \gamma$ . We have that

$$\begin{aligned} \sigma(A) &= \int_{\mathbb{S}^d} d_{\text{Cap}(y,\tilde{\rho}_i)}(A) \, d\sigma(y) \\ &= \int_{\text{Cap}(x,\rho+\tilde{\rho}_i)} d_{\text{Cap}(y,\tilde{\rho}_i)}(A) \, d\sigma(y) \\ &\leq \int_{\text{Cap}(x,\rho)} d_{\text{Cap}(y,\tilde{\rho}_i)}(A) \, d\sigma(y) + \sigma(\text{Cap}_{\rho+\tilde{\rho}_i}) - \sigma(\text{Cap}_\rho). \end{aligned}$$

Since  $\tilde{\rho}_i < \gamma$ , dividing by  $\sigma(\text{Cap}_\rho)$ , we obtain

$$\begin{aligned} \mathbb{E}_{y \in \text{Cap}(x,\rho)} [d_{\text{Cap}(y,\tilde{\rho}_i)}(A)] &\geq \frac{\sigma(A)}{\sigma(\text{Cap}_\rho)} - \frac{\sigma(\text{Cap}_{\rho+\tilde{\rho}_i}) - \sigma(\text{Cap}_\rho)}{\sigma(\text{Cap}_\rho)} \\ &> d_{\text{Cap}(x,\rho)}(A) - \frac{\varepsilon}{4}. \end{aligned}$$

There must then exist  $y_i \in \text{Cap}(x, \rho)$  for which  $d_{\text{Cap}(y_i,\tilde{\rho}_i)}(A) > d_{\text{Cap}(x,\rho)}(A) - \varepsilon/4$ ; fix one such  $y_i$  for each  $1 \leq i \leq N$ , and let  $T_{y_i \rightarrow x_i} \in \text{O}(\mathbb{R}^{d+1})$  be any rotation taking  $y_i$  to  $x_i$  (and thus taking  $\text{Cap}(y_i, \tilde{\rho}_i)$  to  $\text{Cap}(x_i, \tilde{\rho}_i)$ ).

We claim that the set

$$\tilde{A} := \bigcup_{i=1}^N T_{y_i \rightarrow x_i}(A \cap \text{Cap}(y_i, \tilde{\rho}_i))$$

satisfies our requirements. Indeed, we have

$$\begin{aligned} \sigma(\tilde{A}) &= \sum_{i=1}^N \sigma(A \cap \text{Cap}(y_i, \tilde{\rho}_i)) = \sum_{i=1}^N d_{\text{Cap}(y_i,\tilde{\rho}_i)}(A) \cdot \sigma(\text{Cap}_{\tilde{\rho}_i}) \\ &> \sum_{i=1}^N \left( d_{\text{Cap}(x,\rho)}(A) - \frac{\varepsilon}{4} \right) \cdot \left( 1 - \frac{\varepsilon}{4} \right) \sigma(\text{Cap}_{\rho_i}) \\ &\geq \left( d_{\text{Cap}(x,\rho)}(A) - \frac{\varepsilon}{2} \right) \sigma(\mathcal{P}) \\ &> d_{\text{Cap}(x,\rho)}(A) - \varepsilon. \end{aligned}$$

Moreover, since  $\text{diam}(P_j) \leq t_0$  and the caps  $\text{Cap}(x_i, \tilde{\rho}_i)$  are (at least)  $2t_0$ -distant from each other, we see that any copy of  $P_j$  in  $\tilde{A} \subset \bigcup_{i=1}^N \text{Cap}(x_i, \tilde{\rho}_i)$  must be entirely contained in one of the the caps  $\text{Cap}(x_i, \tilde{\rho}_i)$ . But then it should also be contained (after rotation by  $T_{y_i \rightarrow x_i}^{-1}$ ) in  $A \cap \text{Cap}(y_i, \tilde{\rho}_i)$ ; this shows that  $\tilde{A}$  does not contain copies of  $P_j$  for any  $1 \leq j \leq N$ , since  $A$  does not, and we are done.  $\square$

### 3.5. Results on the spherical independence density

We are finally ready to start a more detailed study of the independence density parameter in the spherical setting.

We start by providing a general lower bound on the independence density of several different configurations in terms of their individual independence densities.

**Lemma 18** (Supermultiplicativity). *For all configurations  $P_1, \dots, P_n$  on  $\mathbb{S}^d$ , we have*

$$\mathbf{m}_{\mathbb{S}^d}(P_1, \dots, P_n) \geq \prod_{i=1}^n \mathbf{m}_{\mathbb{S}^d}(P_i).$$

*Proof.* Choose, for each  $1 \leq i \leq n$ , a set  $A_i \subset \mathbb{S}^d$  which avoids configuration  $P_i$ . By taking independent rotations  $R_i A_i$  of each set  $A_i$ , we see that

$$\begin{aligned} \mathbb{E}_{R_1, \dots, R_n \in O(\mathbb{R}^{d+1})} \left[ \sigma \left( \bigcap_{i=1}^n R_i A_i \right) \right] &= \int_{\mathbb{S}^d} \prod_{i=1}^n \mathbb{E}_{R_i \in O(\mathbb{R}^{d+1})} [A_i(R_i^{-1}x)] d\sigma(x) \\ &= \prod_{i=1}^n \sigma(A_i). \end{aligned}$$

There must then exist  $R_1, \dots, R_n \in O(\mathbb{R}^{d+1})$  for which

$$\sigma \left( \bigcap_{i=1}^n R_i A_i \right) \geq \prod_{i=1}^n \sigma(A_i).$$

Since  $\bigcap_{i=1}^n R_i A_i$  avoids all configurations  $P_1, \dots, P_n$  and the sets  $A_1, \dots, A_n$  were chosen arbitrarily, the result follows.  $\square$

Using supersaturation, we can show that this lower bound is essentially tight when the configurations considered are all admissible and each one is at a different size scale. Intuitively, this happens because the constraints of avoiding each of these configurations will act at distinct scales and thus not correlate with each other.

**Theorem 11** (Asymptotic independence). *For every admissible configuration  $P_1, \dots, P_n$  on  $\mathbb{S}^d$  and every  $0 < \varepsilon \leq 1$ , there is a positive increasing function  $f : (0, 1] \rightarrow (0, 1]$  such that the following holds: whenever  $0 < t_1, \dots, t_n \leq 1$  satisfy  $t_{i+1} \leq f(t_i)$  for  $1 \leq i < n$ , we have*

$$\left| \mathbf{m}_{\mathbb{S}^d}(t_1 P_1, \dots, t_n P_n) - \prod_{i=1}^n \mathbf{m}_{\mathbb{S}^d}(t_i P_i) \right| \leq \varepsilon.$$

*Proof.* We have already seen that  $\mathbf{m}_{\mathbb{S}^d}(t_1 P_1, \dots, t_n P_n) \geq \prod_{i=1}^n \mathbf{m}_{\mathbb{S}^d}(t_i P_i)$ , so it suffices to show that  $\mathbf{m}_{\mathbb{S}^d}(t_1 P_1, \dots, t_n P_n) \leq \prod_{i=1}^n \mathbf{m}_{\mathbb{S}^d}(t_i P_i) + \varepsilon$  for suitably separated  $t_1, \dots, t_n \leq 1$ . We will do so by induction on  $n$ , with the base case  $n = 1$  being trivial (and taking  $f \equiv 1$ ).

Suppose then  $n \geq 2$  and we have already proven the result for  $n - 1$  configurations. Let  $\tilde{f} : (0, 1] \rightarrow (0, 1]$  be the function promised by the theorem applied to the  $n - 1$  configurations  $P_2, \dots, P_n$  and with

accuracy  $\varepsilon$ , so that whenever  $0 < t_2 \leq 1$  and  $0 < t_{j+1} \leq \tilde{f}(t_j)$  for each  $2 \leq j < n$ , we have

$$\mathbf{m}_{\mathbb{S}^d}(t_2P_2, \dots, t_nP_n) \leq \prod_{j=2}^n \mathbf{m}_{\mathbb{S}^d}(t_jP_j) + \varepsilon.$$

By the corollary to the Supersaturation Theorem (Corollary 2), for all  $0 < t_1 \leq 1$ , there is  $\delta_0 = \delta_0(\varepsilon; t_1P_1) > 0$  such that

$$\sigma(\mathcal{Z}_{\delta_0}(\varepsilon)[A]) \geq \mathbf{m}_{\mathbb{S}^d}(t_1P_1) + \varepsilon \implies A \text{ contains a copy of } t_1P_1.$$

Applying Lemma 17 with radius  $\rho = \delta_0$ , we see there is  $t_0 = t_0(\varepsilon, \delta_0) > 0$  for which

$$\mathbf{m}_{\text{Cap}(x, \delta_0)}(t_2P_2, \dots, t_nP_n) \leq \mathbf{m}_{\mathbb{S}^d}(t_2P_2, \dots, t_nP_n) + \varepsilon$$

holds whenever  $0 < t_2, \dots, t_n \leq t_0/2$ .

Let now  $0 < t_1, \dots, t_n \leq 1$  be numbers satisfying

$$t_2 \leq t_0(\varepsilon, \delta_0(\varepsilon; t_1P_1))/2 \quad \text{and} \quad t_{j+1} \leq \tilde{f}(t_j) \text{ for all } 2 \leq j < n.$$

If  $A \subset \mathbb{S}^d$  does not contain copies of  $t_1P_1, \dots, t_nP_n$ , then by the preceding discussion, we must have  $\sigma(\mathcal{Z}_{\delta_0}(\varepsilon)[A]) < \mathbf{m}_{\mathbb{S}^d}(t_1P_1) + \varepsilon$  and, for all  $x \in \mathbb{S}^d$ ,

$$\begin{aligned} d_{\text{Cap}(x, \delta_0)}(A) &\leq \mathbf{m}_{\text{Cap}(x, \delta_0)}(t_2P_2, \dots, t_nP_n) \leq \mathbf{m}_{\mathbb{S}^d}(t_2P_2, \dots, t_nP_n) + \varepsilon \\ &\leq \prod_{j=2}^n \mathbf{m}_{\mathbb{S}^d}(t_jP_j) + 2\varepsilon. \end{aligned}$$

This means that, inside caps  $\text{Cap}(x, \delta_0)$  of radius  $\delta_0$ ,  $A$  has density less than  $\varepsilon$  (when  $x \notin \mathcal{Z}_{\delta_0}(\varepsilon)[A]$ ) except on a set of measure at most  $\mathbf{m}_{\mathbb{S}^d}(t_1P_1) + \varepsilon$ , when it instead has density at most  $\prod_{j=2}^n \mathbf{m}_{\mathbb{S}^d}(t_jP_j) + 2\varepsilon$ . Taking averages, we conclude that

$$\begin{aligned} \sigma(A) &= \mathbb{E}_{x \in \mathbb{S}^d} [d_{\text{Cap}(x, \delta)}(A)] \\ &\leq \varepsilon + (\mathbf{m}_{\mathbb{S}^d}(t_1P_1) + \varepsilon) \left( \prod_{j=2}^n \mathbf{m}_{\mathbb{S}^d}(t_jP_j) + 2\varepsilon \right) \\ &\leq 6\varepsilon + \prod_{i=1}^n \mathbf{m}_{\mathbb{S}^d}(t_iP_i). \end{aligned}$$

It thus suffices to take the function  $f : (0, 1] \rightarrow (0, 1]$  given by

$$f(t) = \min \left\{ \tilde{f}(t), \frac{t_0(\varepsilon/6, \delta_0(\varepsilon/6; tP_1))}{2} \right\}$$

to conclude the induction. □

Note that this result provides a partial answer to the analogue of question (Q1) in the spherical setting: if  $P$  is admissible, then  $\mathbf{m}_{\mathbb{S}^d}(t_1P, t_2P, \dots, t_nP)$  decays exponentially with  $n$  as the ratios  $t_{j+1}/t_j$  between consecutive scales go to zero (recall from Lemma 10 that  $\mathbf{m}_{\mathbb{S}^d}(tP)$  is bounded away from both zero and one for  $0 < t \leq 1$ ).

By considering an infinite sequence of ‘counterexamples’ as we did in our proof of Bourgain’s Theorem (Theorem 2), we immediately obtain from Theorem 11 the following result.

**Corollary 3.** *Let  $P \subset \mathbb{S}^d$  be an admissible configuration. If  $A \subseteq \mathbb{S}^d$  has positive measure, then there is some number  $t_0 > 0$  such that  $A$  contains a congruent copy of  $tP$  for all  $t \leq t_0$ .*

This corollary can be seen as the counterpart to Bourgain’s Theorem in the spherical setting, where it is impossible to consider arbitrarily large dilates (the equivalent result of containing all sufficiently small dilates of a configuration in the Euclidean setting also holds with the same proof).

We will next prove that the independence density function  $P \mapsto \mathbf{m}_{\mathbb{S}^d}(P)$  is continuous on the set of admissible configurations on  $\mathbb{S}^d$ . Before doing so, it is interesting to note that a similar result does not hold for two-point configurations on the unit circle  $\mathbb{S}^1$  (which can be seen as the very first instance of nonadmissible configurations). Indeed, it was shown by DeCorte and Pikhurko [7] that  $\mathbf{m}_{\mathbb{S}^1}(\{u, v\})$  is discontinuous at a configuration  $\{u, v\} \subset \mathbb{S}^1$  whenever the arc length between  $u$  and  $v$  is a rational multiple of  $2\pi$  with odd denominator.

**Theorem 12** (Continuity of the independence density). *For any  $n \geq 1$ , the function  $(P_1, \dots, P_n) \mapsto \mathbf{m}_{\mathbb{S}^d}(P_1, \dots, P_n)$  is continuous on the set of  $n$  admissible spherical configurations.*

*Proof.* For simplicity of exposition, we will prove the result in the case of only one forbidden configuration, but the general case follows from the same argument.

Fix some  $\varepsilon > 0$  and some admissible configuration  $P$  on  $\mathbb{S}^d$  and let  $c(\varepsilon) > 0$  be the constant promised by the Supersaturation Theorem (Theorem 10). By equicontinuity (Lemma 15), there exists  $\delta > 0$  such that

$$|I_Q(A) - I_P(A)| \leq \varepsilon \quad \text{for all } Q \in \mathcal{B}(P, \delta), A \subseteq \mathbb{S}^d.$$

Suppose  $Q \in \mathcal{B}(P, \delta)$  and  $A \subset \mathbb{S}^d$  is a measurable set avoiding  $Q$ ; we must then have  $I_P(A) \leq c(\varepsilon)$ , and so  $\sigma(A) \leq \mathbf{m}_{\mathbb{S}^d}(P) + \varepsilon$ . Optimizing over  $A$ , we conclude that  $\mathbf{m}_{\mathbb{S}^d}(Q) \leq \mathbf{m}_{\mathbb{S}^d}(P) + \varepsilon$  whenever  $Q \in \mathcal{B}(P, \delta)$ .

Now, write  $P = \{v_1, \dots, v_k\}$  and consider the function  $g_P : (\mathbb{S}^d)^k \times \mathcal{O}(\mathbb{R}^{d+1}) \rightarrow \mathbb{R}$  given by

$$g_P(x_1, \dots, x_k, T) := \sum_{i=1}^k \|x_i - Tv_i\|.$$

Note that this function is continuous and that  $\min_{T \in \mathcal{O}(\mathbb{R}^{d+1})} g_P(x_1, \dots, x_k, T) = 0$  if and only if  $(x_1, \dots, x_k)$  is congruent to  $(v_1, \dots, v_k)$ .

By inner regularity, we can find a compact set  $A \subset \mathbb{S}^d$  which avoids  $P$  and has measure  $\sigma(A) \geq \mathbf{m}_{\mathbb{S}^d}(P) - \varepsilon$ . The continuous function  $g_P$  attains a minimum on the compact set  $A^k \times \mathcal{O}(\mathbb{R}^{d+1})$ ; denote this minimum by  $\gamma$  and note that  $\gamma > 0$  since  $A$  avoids  $P$ . Let us show that  $A$  also avoids  $Q$ , for all  $Q \in \mathcal{B}(P, \gamma/2k)$ . Indeed, writing  $Q = \{u_1, \dots, u_k\}$  (with the labels chosen so as to minimize their distance to the corresponding points of  $P$ ), for any points  $x_1, \dots, x_k \in A$  and any  $T \in \mathcal{O}(\mathbb{R}^{d+1})$ , we have that

$$\begin{aligned} \sum_{i=1}^k \|x_i - Tu_i\| &\geq \sum_{i=1}^k \left| \|x_i - Tv_i\| - \|Tu_i - Tv_i\| \right| \\ &\geq g_P(x_1, \dots, x_k, T) - k\|Q - P\|_\infty, \end{aligned}$$

which is at least  $\gamma/2 > 0$  if  $\|Q - P\|_\infty \leq \gamma/2k$ . For such configurations, we then obtain

$$\mathbf{m}_{\mathbb{S}^d}(Q) \geq \sigma(A) \geq \mathbf{m}_{\mathbb{S}^d}(P) - \varepsilon.$$

We conclude that  $|\mathbf{m}_{\mathbb{S}^d}(Q) - \mathbf{m}_{\mathbb{S}^d}(P)| \leq \varepsilon$  whenever  $\|Q - P\|_\infty \leq \min\{\delta, \gamma/2k\}$ , finishing the proof. □

As our definition of the independence density  $\mathbf{m}_{\mathbb{S}^d}(P)$  involved a supremum over all  $P$ -avoiding measurable sets  $A \subseteq \mathbb{S}^d$ , it is not immediately clear whether there actually exists a measurable  $P$ -avoiding set attaining this extremal value of density. In fact, such a result is false in the case where  $d = 1$  and we are considering two-point configurations  $\{u, v\} \subset \mathbb{S}^1$ : if the length of the arc between  $u$

and  $v$  is *not* a rational multiple of  $\pi$ , it was shown by Székely [23] that  $\mathbf{m}_{\mathbb{S}^1}(\{u, v\}) = 1/2$ , but there is no  $\{u, v\}$ -avoiding measurable set of density  $1/2$ .

We will now show that extremizer sets exist whenever the configuration we are forbidding is admissible. Note that the result also holds (with essentially unchanged proof) when forbidding several admissible configurations; this generalizes to higher-order configurations a theorem of DeCorte and Pikhurko [7] for forbidden distances on the sphere.

**Theorem 13** (Existence of extremizers). *If  $P \subset \mathbb{S}^d$  is an admissible configuration, then there exists a  $P$ -avoiding measurable set  $A \subseteq \mathbb{S}^d$  attaining  $\sigma(A) = \mathbf{m}_{\mathbb{S}^d}(P)$ .*

*Proof.* Let  $A_1, A_2, \dots \subseteq \mathbb{S}^d$  be a sequence of  $P$ -avoiding measurable sets satisfying  $\lim_{i \rightarrow \infty} \sigma(A_i) = \mathbf{m}_{\mathbb{S}^d}(P)$ . By passing to a subsequence if necessary, we may assume that  $(A_i)_{i \geq 1}$  converges to some function  $A \in \mathcal{B}_\infty$  in the weak\* topology of  $L^\infty(\mathbb{S}^d)$ . We shall prove two things:

- (i) the limit function  $A$  is  $\{0, 1\}$ -valued almost everywhere, so we can identify it with its support  $\text{supp } A$ ;
- (ii) after possibly modifying it on a zero-measure set, this set  $A$  will avoid  $P$ .

With these two results we will be done, since  $\sigma(A) = \lim_{i \rightarrow \infty} \sigma(A_i) = \mathbf{m}_{\mathbb{S}^d}(P)$ .

By weak\* convergence, we know that  $0 \leq A \leq 1$  almost everywhere, and by weak\* continuity, (Lemma 14) we also have  $I_P(A) = \lim_{i \rightarrow \infty} I_P(A_i) = 0$ . From this, we easily conclude that  $I_P(\text{supp } A) = 0$ , and also

$$\sigma(\text{supp } A) = \int_{\mathbb{S}^d} \text{supp } A(x) \, d\sigma(x) \geq \int_{\mathbb{S}^d} A(x) \, d\sigma(x) = \mathbf{m}_{\mathbb{S}^d}(P). \tag{16}$$

But Lemma 11 implies that  $\sigma(\text{supp } A) \leq \mathbf{m}_{\mathbb{S}^d}(P)$ , which by (16) and the fact that  $0 \leq A \leq 1$ , can only happen if  $A = \text{supp } A$  almost everywhere. This proves (i).

Identifying  $A$  with its support and using that  $I_P(A) = 0$ , Lemma 11 implies we can remove a zero-measure subset of  $A$  in order to remove all copies of  $P$ . This proves item (ii) and finishes the proof of the theorem. □

To conclude, let us make explicit what we can say about the possible independence densities when forbidding  $n$  distinct contractions of an admissible configuration  $P$ ; due to lack of dilation invariance in the spherical setting, characterizing these values in terms of simpler quantities is much harder than it is in the Euclidean setting.

Denote  $\mathcal{M}_n^{\mathbb{S}^d}(P) := \{\mathbf{m}_{\mathbb{S}^d}(t_1P, t_2P, \dots, t_nP) : 0 < t_1 < t_2 < \dots < t_n \leq 1\}$ . Due to continuity of  $\mathbf{m}_{\mathbb{S}^d}$  (Theorem 12), this set is an interval and its upper extremity is  $\sup_{0 < t \leq 1} \mathbf{m}_{\mathbb{S}^d}(tP)$ . By supermultiplicativity (Lemma 18), the lower extremity of  $\mathcal{M}_n^{\mathbb{S}^d}(P)$  is at least  $\inf_{0 < t \leq 1} \mathbf{m}_{\mathbb{S}^d}(tP)^n$ , and by asymptotic independence (Theorem 11), it can be at most  $\inf_{0 < t \leq 1} \mathbf{m}_{\mathbb{S}^d}(tP) \cdot \liminf_{t \rightarrow 0} \mathbf{m}_{\mathbb{S}^d}(tP)^{n-1}$ .

#### 4. Concluding remarks and open problems

Our results leave open the question of what happens when the configurations we forbid are *not* admissible. There are two different reasons for a given configuration (either on the space or on the sphere) to not be admissible, so let us examine them separately.

The first reason is that  $P$  is degenerate, meaning that its points are affinely dependent if we are on  $\mathbb{R}^d$  or linearly dependent if we are on  $\mathbb{S}^d$ . In the Euclidean setting, Bourgain [2] showed an example of sets  $A_d \subset \mathbb{R}^d$  (for each  $d \geq 2$ ) which have positive density but which avoid arbitrarily large dilates of a degenerate three-point configuration of the form  $\{-v, 0, v\}$ . These sets then show that the conclusion of Bourgain’s Theorem (and thus also the conclusion of our Theorem 4) is false for this degenerate configuration.

This counterexample was later generalized by Graham [13], who showed that a result like Bourgain’s Theorem *can only hold* if  $P$  is contained on the surface of some sphere of finite radius (as is always the

case when  $P$  is nondegenerate). In fact, Graham's result implies (for instance) that

$$\mathbf{m}_{\mathbb{R}^d}(P, \sqrt{3}P, \sqrt{5}P, \sqrt{7}P, \dots) > 0$$

whenever  $P \subset \mathbb{R}^d$  is nonspherical, that is, not contained on the surface of any sphere. Some kind of nondegeneracy hypothesis is thus necessary both for Bourgain's result and for our Theorem 4.<sup>11</sup>

It is interesting to note, however, that more recent results of Ziegler [24, 25] (generalizing a theorem of Furstenberg, Katznelson and Weiss [12] for three-point configurations) show that every set  $A \subseteq \mathbb{R}^d$  of positive upper density is *arbitrarily close* to containing all large enough dilates of any finite configuration  $P \subset \mathbb{R}^d$ . More precisely, denoting by  $A_\delta$  the set of all points at distance at most  $\delta$  from the set  $A$ , Ziegler proved the following.

**Theorem 14.** *Let  $A \subseteq \mathbb{R}^d$  be a set of positive upper density and  $P \subset \mathbb{R}^d$  be a finite set. Then there exists  $t_0 > 0$  such that, for any  $t \geq t_0$  and any  $\delta > 0$ , the set  $A_\delta$  contains a configuration congruent to  $tP$ .*

The proof of this theorem is ergodic theoretic in nature, making essential use of deep and difficult results regarding nilflows and the characteristic factors of nonconventional ergodic averages. It unfortunately does not seem to follow from our methods.

Let us now turn to the second reason for a configuration  $P$  on  $\mathbb{R}^d$  or  $\mathbb{S}^d$  to be nonadmissible, namely that it contains  $d + 1$  points (if it has more than  $d + 1$  points, then it is obviously degenerate). In this case, we cannot apply the same strategy we used to prove the Counting Lemmas, and it is not clear whether they or the analogues of Bourgain's Theorem are true. We conjecture that they are whenever  $d \geq 2$ , so that we can remove the cardinality condition from the statement of Bourgain's result and of our 'asymptotic independence' Theorem 4 and Theorem 11.

In particular, let us make more explicit the simplest case of this conjecture, which is an obvious question left open since the results of Bourgain and of Furstenberg, Katznelson and Weiss:

**Conjecture 1.** *Let  $A \subset \mathbb{R}^2$  be a set of positive upper density and let  $u, v, w \in \mathbb{R}^2$  be noncollinear points. Then, there exists  $t_0 > 0$  such that for any  $t \geq t_0$ , the set  $A$  contains a configuration congruent to  $\{tu, tv, tw\}$ .*

Another question we ask is related to a suspected compatibility condition between the Euclidean and spherical settings. Since  $\mathbb{S}^d$  resembles  $\mathbb{R}^d$  at small scales, it seems geometrically intuitive that  $\mathbf{m}_{\mathbb{S}^d}(tP)$  should get increasingly close to  $\mathbf{m}_{\mathbb{R}^d}(P)$  as  $t \rightarrow 0$  whenever  $P$  is a contractible configuration on  $\mathbb{S}^d$  (it is easy to show that a configuration  $P \subset \mathbb{S}^d$  is contractible if and only if it is contained in a  $d$ -dimensional affine subspace, so we can embed it in  $\mathbb{R}^d$ ). We ask whether this intuition is indeed correct (i.e., is it true that  $\lim_{t \rightarrow 0} \mathbf{m}_{\mathbb{S}^d}(tP) = \mathbf{m}_{\mathbb{R}^d}(P)$  for all contractible configurations  $P \subset \mathbb{S}^d$ ?).

In a more combinatorial perspective, we wish to know whether an analogue of the Hypergraph Removal Lemma holds for forbidden geometrical configurations. In intuitive terms, the question we ask is whether a measurable set  $A$  (either on  $\mathbb{R}^d$  or on  $\mathbb{S}^d$ ) which contains 'few' copies of some given configuration  $P$  can be made  $P$ -avoiding by removing only 'a few' of its points.<sup>12</sup> Such a result would then explain geometrical sets having few copies of  $P$  as those which are close to a set avoiding this configuration, and it trivially implies the corresponding Supersaturation Theorem; note that this is a quantitative and stronger version of our zero-measure removal Lemmas 2 and 11.

Finally, it would be very interesting to have a way of obtaining good upper bounds for the independence densities of a given configuration or family of configurations. There are several papers (see [1, 6] and the references therein) which consider this question in the case of a single two-point configuration, drawing on powerful methods from the theory of conic optimization and representation theory, and it is already quite challenging in this simplest case. Oliveira and Vallentin also considered the case of several

<sup>11</sup>We believe that the same is true for the spherical analogue of Theorem 4, namely Theorem 11, though we do not know of a counterexample.

<sup>12</sup>On the unit sphere, this property would more formally read: whenever  $A \subseteq \mathbb{S}^d$  satisfies  $I_P(A) \leq \varepsilon$ , there is a subset  $E \subset A$  of measure  $\sigma(E) \leq o_{\varepsilon \rightarrow 0}(1)$  such that  $A \setminus E$  avoids  $P$  (where  $o_{\varepsilon \rightarrow 0}(1)$  denotes a quantity that goes to zero as  $\varepsilon \rightarrow 0$ ). Similarly in the Euclidean setting.



forbidden two-point configurations in Euclidean space [16] and in arbitrary compact, connected, rank-one symmetric spaces [17]; they use linear and semidefinite programming methods to prove that the independence density of  $n$  distinct two-point configurations decays exponentially with  $n$  if their sizes are sufficiently far apart.<sup>13</sup>

We believe that the study of the independence density for higher-order configurations in the optimization setting is also worthwhile, since they serve as model problems for symmetric optimization problems depending on higher-order relations and might prove very fruitful in new methods developed.

**Acknowledgements.** This work was carried out while the author was a Ph.D. student at the University of Cologne. The author would like to thank Fernando de Oliveira Filho, Lucas Slot and Frank Vallentin for many helpful discussions. We also thank the anonymous reviewer, Fernando de Oliveira Filho and Frank Vallentin for several suggestions which improved the presentation of this paper.

**Competing interest.** The author has no competing interest to declare.

**Financial support.** This work was supported by the European Union's EU Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie Actions Grant Agreement No 764759 (MINOA), and by the Dutch Research Council (NWO) as part of the NETWORKS programme (grant no. 024.002.003).

## References

- [1] C. Bachoc, A. Passuello and A. Thiery, 'The density of sets avoiding distance 1 in Euclidean space', *Discrete Comput. Geom.* **53** (2015), 783–808.
- [2] J. Bourgain, 'A Szemerédi type theorem for sets of positive density in  $\mathbb{R}^k$ ', *Israel J. Math.* **54** (1986), 307–316.
- [3] B. Bukh, 'Measurable sets with excluded distances', *Geom. Funct. Anal.* **18** (2008), 668–697.
- [4] D. Castro-Silva, 'Quasirandomness in additive groups and hypergraphs', Preprint, 2021, [arXiv:2107.01463](https://arxiv.org/abs/2107.01463).
- [5] F. Dai and Y. Xu, *Approximation Theory and Harmonic Analysis on Spheres and Balls* (Springer Monographs in Mathematics, Springer, New York, 2013).
- [6] E. DeCorte, F. M. de Oliveira Filho and F. Vallentin, 'Complete positivity and distance-avoiding sets', *Mathematical Programming* (2020).
- [7] E. DeCorte and O. Pikhurko, 'Spherical sets avoiding a prescribed set of angles', *Int. Math. Res. Not. (IMRN)* (2016), 6095–6117.
- [8] C. F. Dunkl, 'Operators and harmonic analysis on the sphere', *Trans. Amer. Math. Soc.* **125** (1966), 250–263.
- [9] P. Erdős, 'Problems and results in combinatorial geometry', *Ann. New York Acad. Sci.* **440** (1985), 1–11.
- [10] P. Erdős and M. Simonovits, 'Supersaturated graphs and hypergraphs', *Combinatorica* **3** (1983), 181–192.
- [11] P. Frankl and R. M. Wilson, 'Intersection theorems with geometric consequences', *Combinatorica* **1** (1981), 357–368.
- [12] H. Furstenberg, Y. Katznelson and B. Weiss, 'Ergodic theory and configurations in sets of positive density', in *Mathematics of Ramsey Theory (Algorithms Combin.)* vol. 5 (Springer, Berlin, 1990), 184–198.
- [13] R. L. Graham, 'Recent trends in Euclidean Ramsey theory', *Discrete Math.* **136** (1994), 119–127.
- [14] P. Mattila, 'Geometry of sets and measures in Euclidean spaces', in *Cambridge Studies in Advanced Mathematics* vol. 44 (Cambridge University Press, Cambridge, 1995).
- [15] R. E. Megginson, *An Introduction to Banach Space Theory* (Graduate Texts in Mathematics) vol. 183 (Springer-Verlag, New York, 1998).
- [16] F. M. de Oliveira Filho and F. Vallentin, 'Fourier analysis, linear programming, and densities of distance avoiding sets in  $\mathbb{R}^n$ ', *J. Eur. Math. Soc. (JEMS)* **12** (2010), 1417–1428.
- [17] F. M. de Oliveira Filho and F. Vallentin, 'A quantitative version of Steinhaus' theorem for compact, connected, rank-one symmetric spaces', *Geom. Dedicata* **167** (2013), 295–307.
- [18] V. Rödl and M. Schacht, 'Regularity lemmas for graphs', in *Fete of Combinatorics and Computer Science* (Bolyai Soc. Math. Stud.) vol. 20 (János Bolyai Math. Soc., Budapest, 2010), 287–325.
- [19] E. M. Stein, *Harmonic Analysis: Real-variable Methods, Orthogonality, and Oscillatory Integrals* (Princeton Mathematical Series) vol. 43 (Princeton University Press, Princeton, NJ, 1993). With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III.
- [20] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces* (Princeton Mathematical Series) no. 32 (Princeton University Press, Princeton, N.J., 1971).
- [21] G. Szegő, *Orthogonal Polynomials* (American Mathematical Society Colloquium Publications) vol. XXIII, fourth edn (American Mathematical Society, Providence, RI, 1975).

<sup>13</sup>This holds only if the real dimension of the considered space is 2 or higher. Oliveira and Vallentin [17, Section 2] provide counterexamples to the corresponding result in spaces of real dimension 1.



- [22] L. A. Székely, 'Remarks on the chromatic number of geometric graphs', in *Graphs and Other Combinatorial Topics* vol. 59 (Teubner-Texte Math. , Teubner, Leipzig, 1983), 312–315.
- [23] L. A. Székely, 'Measurable chromatic number of geometric graphs and sets without some distances in Euclidean space', *Combinatorica* **4** (1984), 213–218.
- [24] T. Ziegler, 'An application of ergodic theory to a problem in geometric Ramsey theory', *Israel J. Math.* **114** (1999), 271–288.
- [25] T. Ziegler, 'Nilfactors of  $\mathbb{R}^m$ -actions and configurations in sets of positive upper density in  $\mathbb{R}^m$ ', *J. Anal. Math.* **99** (2006), 249–266.